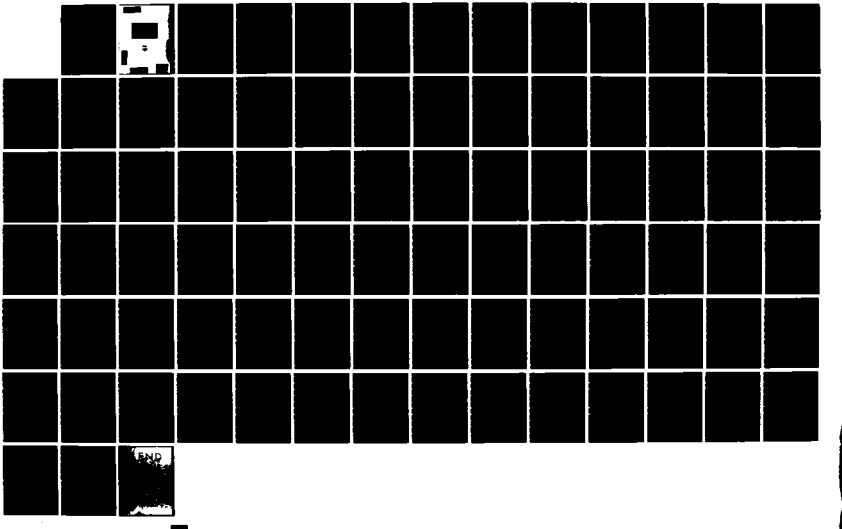
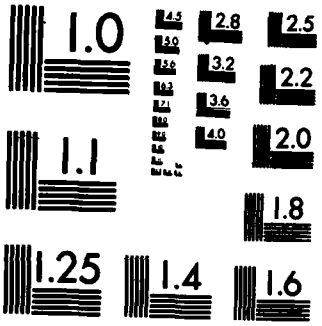
AD-A147 378 AMBIGUITY AND UNCERTAINTY IN PROBABILISTIC INFERENCE 1/1
(U) CHICAGO UNIV IL CENTER FOR DECISION RESEARCH
H J EINHORN ET AL. JUN 84 TR-10 N00014-84-C-0018
UNCLASSIFIED F/G 12/1 NL

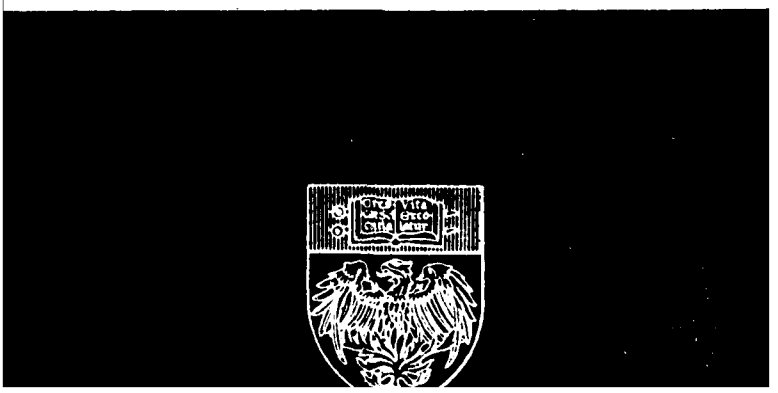



1.0
1.1
1.25
1.4
1.6
1.8
2.0
2.2
2.5
2.8
3.2
3.6
4.0

AMBIGUITY AND UNCERTAINTY IN PROBABILISTIC INFERENCE

Hillel J. Einhorn Robin M. Hogarth
University of Chicago
Graduate School of Business
Center for Decision Research

June 1984



(1)

AMBIGUITY AND UNCERTAINTY IN PROBABILISTIC INFERENCE

Hillel J. Einhorn Robin M. Hogarth
University of Chicago
Graduate School of Business
Center for Decision Research

June 1984

Sponsored by:
Engineering Psychology Programs
Office of Naval Research
Contract Number, N00014-84-C-0018
Work Unit Number, NR 197-080

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

Abbreviated title: Ambiguity and Uncertainty

DTIC
ELECTE
JAN 13 1984

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 10	2. GOVT ACCESSION NO. AD A147328	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Ambiguity and Uncertainty in Probabilistic Inference		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Hillel J. Einhorn and Robin M. Hogarth		8. CONTRACT OR GRANT NUMBER(s) N00014-84-C-0018
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Decision Research, Graduate School of Business, University of Chicago, 1101 East 58th Street, Chicago, Illinois 60637		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 197-080
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 North Quincy Street Arlington, VA 22217		12. REPORT DATE June 1984 13. NUMBER OF PAGES 68
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Ambiguity; second-order uncertainty; probabilistic inference; risky choice.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Ambiguity results from having limited knowledge of the process that generates outcomes. It is argued that many real-world processes are perceived to be ambiguous; moreover, as Ellsberg (1961) demonstrated, this poses problems for theories of probability operationalized via choices amongst gambles. A descriptive model of how people make judgments under

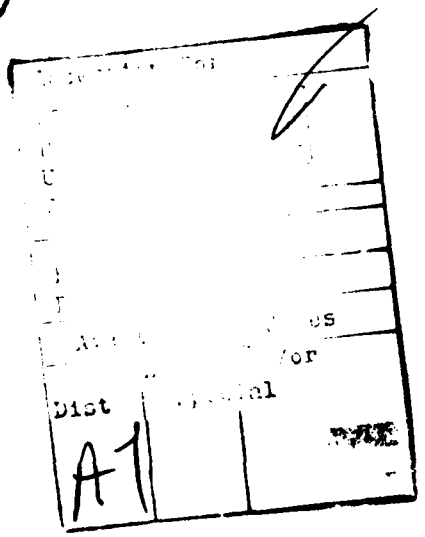
DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ambiguity is proposed. The model assumes an anchoring-and-adjustment process in which an initial estimate provides the anchor, and adjustments are made for "what might be." The latter is modeled as the result of a mental simulation process where the size of the simulation is a function of the amount of ambiguity, and differential weighting of imagined probabilities reflects one's attitude toward ambiguity. A two-parameter model of this process is shown to be consistent with: Ellsberg's original paradox, the non-additivity of complementary probabilities, current psychological theories of risk, and Keynes' idea of the "weight of evidence." The model is tested in four experiments involving both individual and group analyses. In experiments 1 and 2, the model is shown to predict judgments quite well; in experiment 3, the inference model is shown to predict choices between gambles; experiment 4 shows how buying and selling prices for insurance are systematically influenced by one's attitude toward ambiguity. The results and model are then discussed with respect to: (1) the importance of ambiguity in assessing uncertainty; (2) the use of cognitive strategies in judgments under ambiguity; (3) the role of ambiguity in risky choice; and, (4) extensions of the model.



AMBIGUITY AND UNCERTAINTY IN PROBABILISTIC INFERENCE

The literature on how people make judgments under uncertainty is large, complex, and rife with controversy (see e.g., Edwards, 1954, 1968; Peterson & Beach, 1967; Slovic & Lichtenstein, 1971; Rappoport & Wallsten, 1972; Slovic, Lichtenstein, & Fischhoff, 1977; Einhorn & Hogarth, 1981; Kahneman, Slovic, & Tversky, 1982; Cohen, 1982; Kyburg, 1983). One reason for the controversy is that while there is agreement that "uncertainty" is a crucial factor in inference, there is much less agreement about its meaning and measurement (*cf.* Tversky & Kahneman, 1982). In particular, while most psychological work on inference has been guided by a Bayesian or subjectivist view of probability, increasing concerns have been expressed about this position (e.g., Cohen, 1977; Shafer, 1978). Central to the Bayesian view is the idea that probability, which is a measure of one's degree of belief, can be operationalized via choices amongst gambles (Savage, 1954). Thus, if two gambles have identical payoffs but one is preferred to the other, it follows that the probability of winning is greater for the chosen alternative.

The subjectivist view of probability gains much of its force by making expressions of uncertainty operational via choices amongst gambles. However, whereas probability is thereby defined precisely, does this procedure capture the essential psychological aspects of uncertainty? In particular, how valid is the assumption that expressions of uncertainty can be captured through choices amongst gambles? An important and direct attack on this assumption was put forward by Daniel Ellsberg (1961) and we examine his arguments below. In doing so, however, we stress that our intent is to understand the psychological bases of uncertainty rather than to critique the normative status of the Bayesian position.

Ellsberg (1961) used the following example to show that the uncertainty people experience has several aspects, one of which is not captured in the usual betting paradigm: Imagine two urns, each containing red and black balls. In urn 1, there are 100 balls but the proportions of red and black are unknown; urn 2 contains 50 red and 50 black balls. Now consider a gamble such that, if you bet on red and it is drawn from the urn you get $100; similarly for black. However, if you bet on the wrong color, the payoff is $0. Imagine having to decide which color to bet on if a ball is to be drawn from urn 1; i.e., the choices are red (R_1), black (B_1), or indifference (I). What about the same choices in urn 2; (R_2), (B_2), or (I)? Most people are indifferent in both cases, suggesting that the subjective probability of red in urn 1 is the same as the known proportion in urn 2--namely .5. However, would you be indifferent to betting on red if urn 1 were to be used vs. betting on red using urn 2 (R_1 vs. R_2)? Similarly, what about B_1 vs. B_2? Many people find that they prefer R_2 over R_1 even though their indifference judgments within both urns imply that, $p(R_1) = p(R_2) = .5$. Furthermore, the same person who prefers R_2 over R_1 may also prefer B_2 over B_1. This pattern of responses is inconsistent with the idea that even a rank order of probabilities can be inferred from choices. Thus, if R_2 is preferred over R_1, this implies that $p(R_2) > p(R_1)$. Moreover, since red and black are complementary events, this means that $p(B_2) < p(B_1)$. However, if B_2 is preferred over B_1, then $p(B_2) > p(B_1)$, which contradicts the preceding inequality. It is also important to note that if $p(R_2) > p(R_1)$ and $p(B_2) > p(B_1)$, then either urn 2 has complementary probabilities summing to more than 1 (super-additivity), or, urn 1 has complementary probabilities summing to less than 1 (sub-additivity). Although Ellsberg did not specifically discuss the non-additivity of complementary probabilities (cf. Fellner, 1961), we shall show that it is intimately

related to the effects of different types of uncertainty on probabilistic judgments.

From our perspective, the importance of Ellsberg's paradox lies in the difference in the nature of the uncertainty between urns 1 and 2. In urn 1, whereas one's best estimate of the proportion may be .5, confidence in that estimate is low. In urn 2, on the other hand, one is at least certain about the uncertainty in the urn. While it may seem strange, and even awkward, to speak of uncertainty as being more or less certain itself, such a concept captures an important aspect of how people make inferences from unknown, or only partially known, generating processes. Indeed, the idea of uncertainty about uncertainty has been considered from time-to-time under the rubrics, "second-order" uncertainty and probabilities for probabilities (e.g., Marschak, 1975). However, whereas this concept has received little support amongst subjectivist statisticians (see e.g., de Finetti, 1977), its status as a psychological factor of importance for understanding choice and inference has been demonstrated experimentally (Becker & Brownson, 1964; Yates & Zukowski, 1976). On the other hand, the process by which such second-order uncertainty is used in inference and the factors that affect its use, have not been systematically studied. To be sure, Ellsberg suggested a number of variables that should affect the "ambiguity" of a situation, including the amount, type, reliability, and degree of conflict in the available information. Indeed, he stated,

> Ambiguity is a subjective variable, but it should be possible to identify 'objectively' some situations likely to present high ambiguity, by noting situations where available information is scanty or obviously unreliable or highly conflicting; or where expressed expectations of different individuals differ widely; or where expressed confidence in estimates tends to be low. Thus, as compared with the effects of familiar production decisions or well-known random processes (like coin-flipping or roulette), the results of Research and Development, or the performance of a new President, or the tactics of an unfamiliar

> opponent are all likely to appear ambiguous. (1961, pp. 660-661).

To specify the concept of ambiguity more precisely, reconsider the urn where the proportion of red and black balls is unknown. From a Bayesian perspective, this situation can be thought of as one in which the judge has a diffuse prior over all possible values of the proportion, p(R). However, imagine that one sampled four balls (without replacement) and got 3 red and 1 black. Note that this result rules out certain values of p(R) and could change one's assessment of other values of p(R). Furthermore, as the sample size increases, one should become more sure as to the actual value of p(R). Therefore, as information increases, ignorance (a uniform distribution), gives way to ambiguity (a non-uniform distribution over all outcomes), which then reduces to a known p(R). However, while it is tempting to equate ambiguity with some statistical measure of the dispersion of the subjective distribution, this is unsatisfactory for the following reason: consider an urn that contains either all red or all black balls but you don't know which. In such a case we can characterize the distribution over p(R) as having half its mass at zero and half at one. Note that the variance or range of this distribution is high, yet, ambiguity is low. The reason is that such a distribution rules out all values of p(R) other than 0 or 1 and is thus close to the case where ambiguity doesn't exist (as in urn 2). Therefore, in accord with its dictionary definition, "having two or more possible meanings," ambiguity is a function of the number of alternative parameter values that are not ruled out (or made implausible) by one's knowledge of the situation. Note that this definition is similar to, but not identical with, statistical measures such as variance, range, and the like.

It is important to note that sample size is only one factor that influences ambiguity since other information can affect the probability

distribution over the parameter of a stochastic process. Thus, imagine an urn factory where employees color balls by throwing them at two adjacent cans of black and red paint from a distance of 20 feet. Given our knowledge of this process, it seems fair to expect that an urn of 100 balls would not contain extreme proportions of red or black. A second example, due to Gardenfors and Sahlin (1982), is particularly illuminating on this issue:

> . . . consider Miss Julie who is invited to bet on the outcome of three different tennis matches. As regards match A, she is very well-informed about the two players . . . . Miss Julie predicts that it will be a very even match and a mere chance will determine the winner. In match B, she knows nothing whatsoever about the relative strength of the contestants . . . and has no other information that is relevant for predicting the winner of the match. Match C is similar to match B except that Miss Julie has happened to hear that one of the contestants is an excellent tennis player, although she does not know anything about which player it is, and that the second player is indeed an amateur so that everybody considers the outcome of the match a foregone conclusion. (pp. 361-362).

Note that the amount and type of information in the three situations is quite different, as is the amount of ambiguity (we would argue that match A has the least ambiguity and match B the most). From our perspective, how does the amount and type of ambiguity affect judgments of the probability of winning or losing the match? Would Miss Julie, for example, judge that each player in the three matches has a .5 chance of winning (or losing)?

Our discussion so far has implied that ambiguity is generally avoided since it adds to the total uncertainty of a situation. Indeed, this is explicitly mentioned by Ellsberg (1961, p. 666) in discussing why new technologies will be resisted more than one would expect on the basis of their first-order probabilities. However, this picture is not completely accurate, as is made clear by another Ellsberg example (as quoted in Becker & Brownson, 1964, pp. 63-4, footnote 4): consider two urns with 1000 balls each. In urn I, each ball is numbered from 1 to 1000 and the probability of drawing any

number is .001. In urn II, there are an unknown number of balls bearing any single number. Thus, there may be 1000 balls with number 687, no balls with this number, or anything in between. If there is a prize for drawing number 687 from the urn, would you prefer to draw from urn I or urn II? Note that urn I has no ambiguity and each numbered ball has the same .001 chance of being drawn. Urn II, on the other hand, can be characterized as inducing extreme ambiguity (i.e., ignorance). However, for many people, the drawing from urn II seems considerably more attractive than from urn I, thereby implying that there are situations in which ambiguity is preferred rather than avoided. This is considered in detail later, but we note here that accounting for such shifts is an important criterion for judging the adequacy of any theory of inference under ambiguity.

Finally, the concepts of ambiguity, second-order uncertainty, and the like, have been of concern in theories of inference quite apart from their role in affecting choice. For example, work on fuzzy sets (Zadeh, 1978), Shafer's theory of evidence (1976), Cohen's (1977) attempt to formalize uncertainty in legal settings, and the elicitation of probability ranges (Wallsten, Forsyth, & Budescu, 1983), all contain ideas concerning the vagueness that can underly probabilities. Indeed, statisticians have provided axiomatic systems for trying to formalize probability ranges and rank orders rather than specific values (e.g., Koopman, 1940). Moreover, early work by Keynes (1921) also addressed the notion of ambiguity by distinguishing between probability and what he called the "weight of evidence." Keynes stated:

> The magnitude of the probability. . .depends upon a balance between what may be termed the favourable and the unfavourable evidence; a new piece of evidence which leaves this balance unchanged, also leaves the probability of the argument unchanged. But it seems that there may be another respect in which some kind of quantitative comparison between arguments is possible. This comparison turns upon a balance, not between the favourable and unfavourable evidence, but between the <u>absolute</u> amounts of

> relevant knowledge and of relevant ignorance respectively. (Keynes, 1921, p. 71, original emphasis).

Plan of the Paper

We first present a descriptive model of how people make probability judgments and choices under varying amounts of ambiguity. We require that our model be able to: (1) Explain the pattern of choices elicited by Ellsberg's problems. This, in turn, implies that the model account for sub-and super-additivity of the probabilities of complementary events; (2) specify the conditions under which people will seek as well as avoid ambiguity; (3) allow for individual differences; and (4) be empirically testable. To meet e criteria, the model is tested in four experiments at both the aggre and individual subject levels. Two of the experiments concern inferenc t s, and two involve choices. The implications of the theory and empirical work are then discussed in relation to: (a) the importance of ambiguity in assessing perceived uncertainty; (b) the use of cognitive strategies in understanding probabilistic judgments under ambiguity; (c) the role of ambiguity in risky choice; and (d) extensions of the model to multiple sources and time periods.

A Descriptive Model

Our model postulates an "anchoring and adjustment" strategy for assessing probabilities. This involves an initial estimate, denoted p_A, and an adjustment to reflect the ambiguity in the situation. Thus, the ensuing judgment, $S(p_A)$, is given by,

$$S(p_A) = p_A + k \qquad (1)$$

where k is the net effect of the adjustment process. To model the adjustment process, we propose that people engage in a mental simulation in which

other values of p are considered by imagining how well they express one's uncertainty. These simulated values are then incorporated into the adjustment term. The rationale for the simulation is that in ambiguous situations, p can be any one of a number of values. By incorporating the range of possible values of p into their judgments, people can maintain sensitivity to both uncertainty and ambiguity.

We further argue that k, the net effect of the simulation, will be affected by three factors: (1) The level of p_A; that is, since $S(p_A)$ varies between 0 and 1, equation (1) implies that $-p_A \leq k \leq (1-p_A)$. This means that the direction of the adjustment must be due, in part, to the value of p_A. Indeed, when $p_A = 0$, $k \geq 0$, and the adjustment (if there is one) must be upward; when $p_A = 1$, $k \leq 0$, so that the adjustment must be downward; when $p_A \neq 0, 1$, adjustments can be up or down; (2) the amount of ambiguity perceived in the situation. We denote this by a parameter θ, which determines the absolute size of the adjustment, i.e., the larger θ, the larger the adjustment; (3) the person's attitude toward ambiguity in the circumstances. This is reflected in the tendency to give differential attention or weight to values of p that are greater or smaller than the initial estimate, p_A. One's attitude toward ambiguity is denoted by β, and this parameter, together with p_A, determines the sign of the net effect of the adjustment, i.e., when k is positive or negative.

To model the adjustment process algebraically, let

$$k = k_g - k_s \qquad (2)$$

where k_g denotes the effect of imagining values of p greater than the initial estimate, and k_s the effect of imagining smaller values. How does perceived ambiguity affects these quantities? To answer this, consider Figure 1, which shows the position of p_A relative to the end points of 0 and 1.

Insert Figure 1 about here

Note that the maximum upward adjustment to p_A is $(1 - p_A)$ and the maximum downward adjustment is p_A. If we restrict the range of θ to the unit interval $(0 \leq \theta \leq 1)$, then maximum adjustments would occur under complete ambiguity $(\theta = 1)$, and zero adjustments under no ambiguity $(\theta = 0)$. This suggests that the effects of simulating values greater and smaller than p_A (k_g and k_s, respectively), can be represented as proportions of the maximum adjustments where θ is the constant of proportionality, i.e.,

$$k_g = \theta(1-p_A) \tag{3a}$$

and

$$k_s = \theta p_A \tag{3b}$$

The development so far ignores the possibility that greater and smaller values than p_A could be differentially weighted. For example, in estimating the chance of an accident in a new technology (high ambiguity), one may start with the estimate offered by the engineering department and then weight larger values of p(accident) more than smaller ones. To account for differential weighting effects, we need only weight either k_g or k_s to affect k. For convenience, we weight k_s (rather than k_g) by β as follows,

$$k_s = \theta p_A^{\beta} \quad (\beta > 0) \tag{4}$$

Thus, the net effect of the adjustment process is given by,

$$\begin{aligned} k &= k_g - k_s \\ &= \theta\left(1-p_A-p_A^{\beta}\right) . \end{aligned} \tag{5}$$

When (5) is substituted into (1), the full model becomes,

$$S(p_A) = p_A + \theta\left(1-p_A-p_A^{\beta}\right) \tag{6}$$

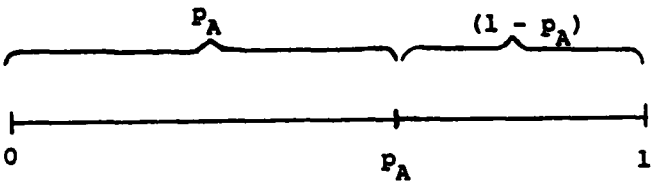


Figure 1. Anchor of p_A and range of adjustments.

We make several points with respect to equation (6). First, note that θ affects the absolute size of the adjustment factor. That is, when there is no ambiguity, $\theta = 0$ and $S(p_A) = p_A$. Thus, θ can be thought of as having a magnifying or dampening effect on one's attitude toward ambiguity in the circumstances, (β). For example, if perceived ambiguity is small, the tendency to weight differentially values of p above and below p_A is of little consequence.

Second, $S(p)$ is regressive with respect to p. This can be illustrated by considering the effects of different values of β in equation (6). Specifically, the three panels of Figure 2 illustrate "ambiguity functions" with $\beta < 1$, $\beta > 1$, and $\beta = 1$. (θ is shown to be the same in all three cases.) It is important to note that each value of β defines a unique

Insert Figure 2 about here

"cross-over" point, p_c, where $S(p) = p$. Thus, in Figure 2a, β defines p_{c1} such that small probabilities are overweighted and larger probabilities underweighted. This form of the function results because $\beta < 1$ implies that more weight is given to smaller values of p rather than larger ones. Therefore, $k < 0$ over most of the range of p. However, when $p_A < p_{c1}$, there are few smaller values of p to consider relative to larger ones. Thus, even when smaller values are weighted more heavily than larger ones, there are more of the latter and $k > 0$. Conversely, when $\beta > 1$, as shown in Figure 2b, $S(p) > p$ over most of the range of p since more weight is given to larger as opposed to smaller p's. However, when $p_A > p_{c2}$, $S(p) < p$ since there are few larger values and $k < 0$. Finally, note that in Figure 2c, when $\beta = 1$, the cross-over point is at .5.

Third, equation (6) implies the conditions under which probability judgments of complementary events are additive (sum to one). Specifically,

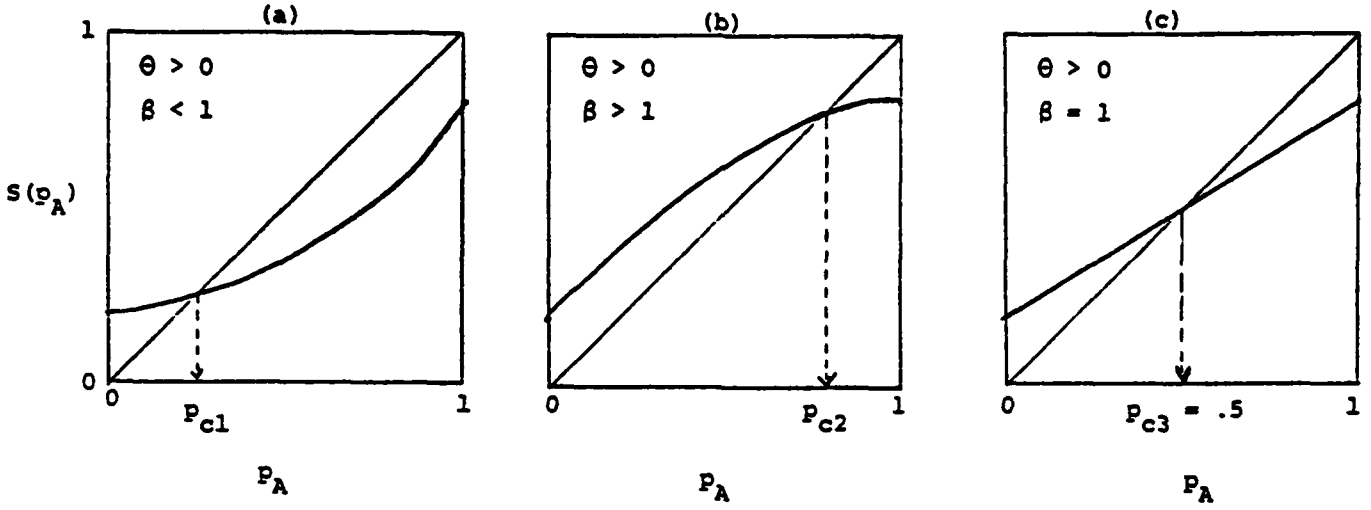


Figure 2. $S(p_A)$ as a function of p_A

consider the sum of $S(p_A)$ and $S(1-p_A)$. This is,

$$S(p_A) + S(1-p_A) = p_A + \theta(1-p_A-p_A^{\beta}) + (1-p_A) + \theta[1-(1-p_A) - (1-p_A)^{\beta}]$$
$$= 1 + \theta[1-p_A^{\beta} - (1-p_A)^{\beta}] \quad (7)$$

Thus, complementary probabilities are additive if: $\theta = 0$, or $\beta = 1$, or $p_A = 0, 1$; otherwise, there is sub-additivity if $\beta < 1$, and super-additivity if $\beta > 1$.

Fourth, there are many ways we could have chosen to incorporate the β parameter in the model. However, not all forms have the same implications, particularly with respect to the additivity of complementary probabilities. We consider several alternative models in Appendix A.

To summarize, the model has two parameters and both are functions of individual and situational factors. The θ parameter reflects perceived ambiguity and the degree to which one simulates values of p that "might be." However, situational factors are also likely to affect θ (across people); e.g., the absolute amount of evidence available, the unreliability of sources, lack of causal knowledge regarding the process generating outcomes, and so on. The β parameter reflects the extent to which one differentially weights in imagination possible values of p that are smaller vs. larger than p_A. As such, β may be related to an optimism-pessimism attitude at the individual level. However, we argue that β will also be influenced by situational variables such as the sign and size of the payoffs that are contingent on the ambiguous probability. For example, if the general effect of ambiguity is to induce caution rather than riskiness, the prospect of an undesirable outcome (e.g., monetary losses) would induce people to pay more attention in imagination to values of p(Loss) that are larger than p_A; similarly, the prospect of a gain would focus attention on smaller values of

p(Gain). We consider this issue further in connection with experiments 3 and 4.

We now consider how the model in (6) explains Ellsberg's original results. Note Figure 2a, where $\theta > 0$ and $\beta < 1$. A person with parameter values in these ranges will "underweight" all p_A above p_C, and "overweight" $p_A < p_C$. This particular pattern explains why most people in Ellsberg's urn example avoid the ambiguous urn 1; that is, $S(p_A = .5) < .50$. However, note that if p_A is less than p_C, $S(p_A) > p_A$ and one would expect the same person who avoided the ambiguous urn when $p_A = .5$, to <u>prefer</u> the ambiguous urn when p_A is sufficiently low (e.g., when $p_A = .001$). The pattern of overweighting small p_A and underweighting moderate-to-large p_A also accounts for some otherwise puzzling results of Goldsmith and Sahlin (as reported in Gardenfors & Sahlin, 1982). They presented subjects with descriptions of either well-known events (e.g., drawing cards from a standard deck), or events about which the subjects had little knowledge (e.g., the likelihood of a bus strike in Verona, Italy next week). Subjects estimated the probabilities of the events and the perceived reliability of their probability estimates. Events with equal probabilities but unequal reliabilities were then used in a lottery set-up. The authors report that,

> . . . <u>for probabilities other than fairly low ones</u>, lottery tickets involving more reliable probability estimates tend to be preferred. (Gärdenfors & Sahlin, 1982, p. 363, our emphasis.)

While the pattern shown in Figure 2a accounts for much data, it does not explain why some people in the Ellsberg task prefer to bet on drawing from the ambiguous urn when $p_A = .5$. However, consider a person with an $S(p_A)$ function as shown in Figure 2b. When $\theta > 0$ and $\beta > 1$, one gets "ambiguity preference" over most of the range of p_A. Thus, when $p_A < p_C$, $S(p_A) > p_A$ and overweighting occurs; when $p_A > p_C$, $S(p_A) < p_A$ and underweighting occurs. Since individual differences are rarely accounted for in research on

decisions under uncertainty, our model has the distinct advantage of positing a general psychological process while allowing for individual differences via particular parameter values. Indeed, this is nicely illustrated by considering people who are indifferent between gambles from ambiguous and unambiguous urns when $p_A = .5$ (as in the Ellsberg case). Our model suggests two distinct types: those for whom $\theta = 0$, and thus, $S(p_A) = p_A$; and those for whom $\theta > 0$ and $\beta = 1$ (shown in Figure 2c). This latter group does not adjust at $p_A = .5$, but does adjust at all other values. Therefore, people characterized by these parameter values will only be indifferent between lotteries at .5.

Finally, we note that our model is relevant to a major psychological theory of risk; namely, "prospect theory" (Kahneman & Tversky, 1979). From our perspective, the treatment of uncertainty in prospect theory is consistent with our approach since a decision-weight function is posited that is remarkably similar to the $S(p_A)$ function shown in Figure 2a. This is not a coincidence since, as Kahneman and Tversky specifically point out, decision weights can be affected by ambiguity. Indeed, they state,

> The decision weight associated with an event will depend primarily on the perceived likelihood of that event, which could be subject to major biases. In addition, decision weights may be affected by other considerations, such as ambiguity or vagueness. Indeed, the work of Ellsberg and Fellner implies that vagueness reduces decision weights. (p. 289)

While our equation (6) could be made fully compatible with the decision-weight function of prospect theory (by restricting its applicability to $0 < p < 1$ and thereby not defining the end points),[1] we wish to emphasize that (6) expresses a <u>class</u> of functions. Therefore, while the decision-weight function of prospect theory expresses a general tendency to treat uncertainty in a particular way, (6) allows for both situational variables and individual differences in the handling of uncertainty.

EXPERIMENTAL TESTS OF THE MODEL

To test our model empirically, we employed two tasks that focused on inference (experiments 1-2) and two dealing with choice (experiments 3-4). In the inference task, people were asked to make probability judgments on the basis of numbers of reports from a source. In experiment 1, we examined the various implications of equation (6). In experiment 2, we used different scenarios to manipulate θ in both a between and within-subjects design. In addition, the consistency of individual differences in strategy (as measured by a person's θ and β parameters) was also considered. Experiment 3 involved an attempt to answer the question: Can an individual's choices between gambles be predicted from knowledge of his or her θ and β parameters obtained from a separate inference task? Finally, in experiment 4, people were asked to be either buyers or sellers of insurance in ambiguous and non-ambiguous situations. Differences between buying and selling prices were then investigated as a function of assumed differences in β parameters. Since experiments 1-3 are all based on the same type of inference task, we first explicate the underlying nature of this task, noting how it differs from other probabilistic tasks considered in the literature.

A Model for Studying Ambiguity in Inference

The prototypical inference that we consider involves a judge assessing the likelihood of the occurrence of an event based on reports received from a source of limited reliability. The task can be thought of as having the elements schematically represented in Figure 3. (1) An event occurs; (2) The event is "sensed" by observers (e.g., witnesses to an accident) who,

Insert Figure 3 about here

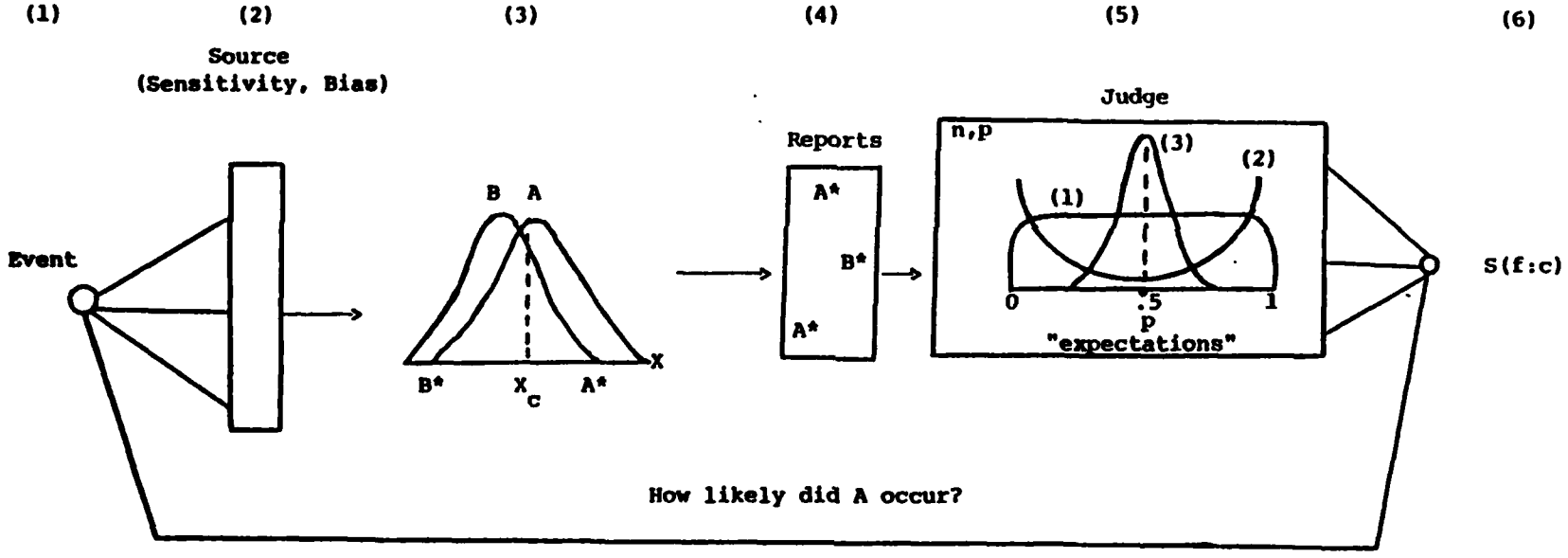


Figure 3. Structure of the inference task

in principle, can be characterized by levels of sensitivity and bias. However, it is important to emphasize that these levels are unknown to the judge (see 5 below); (3) The observers report what they saw. We denote A* as the report of event A, and B* as the report of event B, where the decision rule is to report A* if the observation is above some critical value X_c, and B* otherwise. The reports can therefore be conceptualized as coming from a signal-detection task; (4) Since there are n observers, n reports are collected. Thus, the n reports can be thought of as the outcomes of n observers reporting on a single trial of a signal detection task. Furthermore, since we do not differentiate between the n observers, we refer to them as coming from a single source; (5) The judge receives the information in the form of f reports for a hypothesis (i.e., f reports of A*) and c reports of an alternative (i.e., c reports of B*), where f+c = n, and p = f/n. The content of the scenario, however, is assumed to give the judge some information as to what values of p to expect in a sample of size n. Specifically, we argue that expectations concerning p will be influenced by, (a) the dissimilarity between events A and B; and (b) the credibility of the source. By "credibility" we mean the sensitivity and response bias of the observers in judging the particular events of interest. For example, imagine that you are a detective investigating a bank robbery where two witnesses claim that the robber has blond hair and one witness claims it is brown. How likely does the robber have blond hair? While the detective knows neither the hit and false alarm rates of the witnesses, nor their response bias for saying "blond" vs. "brown," he may know something about the quality of eye-witnesses in a robbery, the confusability of blond and brown in the circumstances, and perhaps something about the motivation of the witnesses. Now contrast this situation where the source is two color television cameras that were filming

the robbery at the bank. Whereas in the former case the detective would expect the reports to conflict (i.e., $0 < p < 1$), in the latter it would be surprising if p were not equal to either 0 or 1.

Note that in Figure 3, we have represented the judge's expectations by three different distributions. In distribution (1), the information about the credibility of the source, the dissimilarity of the signals, and the size of the sample, does not rule out many values of p. This is a highly ambiguous situation and would, for example, characterize the detective trying to judge evidence from witnesses. Distribution (2) characterizes expectations based on a highly credible source that discriminates between dissimilar signals; e.g., evidence from cameras filming the robbery. We believe that ambiguity is low here since our knowledge of the process that generates evidence rules out most values of p. Distribution (3) also represents a situation of low ambiguity, but it is quite different from (2). Indeed, (3) is likely to result when the credibility of the source is particularly low and/or the signals are very similar, in direct opposition to the conditions that produce (2). For example, imagine a taste-test between Pepsi vs. Coke for randomly chosen shoppers. If we believe that the two drinks have a very similar taste and that most shoppers are not able to tell the difference, we would expect the proportion of reports for either product to be around .5. Thus, results from such a test might be seen as most closely resembling the drawing of balls from an urn with known $p = .5$. It is interesting to note that whereas some authors have equated increased reliability of evidence with less ambiguity (as suggested by Ellsberg, for example), distribution (3) shows that <u>decreased</u> reliability can also lead to low ambiguity. Another way to express this is to note that high reliability implies low ambiguity (distribution (2)), but low ambiguity does not imply high reliability (since distribution (3) could be

involved); (6) The judge combines the information from the reports with expectations about p to reach an assessment of the likelihood of A.

The structure of this task is both similar to and different from several probabilistic models of the inference process. First, it is similar to cascaded inference in that the judge is making inferences about an event on the basis of unreliable reports (cf. Schum & Kelley, 1973; Schum, 1980). However, in contrast to studies of cascaded inference, the judge does not know the precise value of the source's reliability; rather, there is ambiguity concerning what this is.

Second, since each observer can be thought of as participating in the same signal detection task, the reports not only reflect their sensitivity to competing signals, but also their bias due to differential payoffs. However, as recently emphasized by Birnbaum (1983), the manner in which the judge treats the observer reports depends on some theory about the observers. For example, the observer reports could be responsive to the prior probabilities of A and B as well as to differential payoffs. We emphasize that in our task the judge is not given precise information about these matters. Furthermore, since the judge only receives information on a single trial, the observers' hit-rate and false-alarm rate are not known. Instead, the observed p, and the judge's expectations about p, become cues to the likelihood that the event occurred.

Third, one might consider our situation as a conventional Bayesian revision task (cf. Edwards, 1968). However, the explicit probabilities necessary to assess the likelihood functions are not provided; and, no base-rate data or prior probabilities are stated. It would, of course, be possible to provide the judge with explicit prior probabilities. This would, however, be extending our paradigm to one where multiple sources of information need to

be combined, i.e., base-rates and individuating information. For the sake of simplicity, we only consider the effects of ambiguity on inferences from a single source and thus do not discuss the effects of explicit base rates (extensions of our model to multiple sources is considered in the Discussion section).

In this model, note that we have explicitly recognized three sources of ambiguity, viz: (a) the dissimilarity between events A and B; (b) the credibility of the source; and (c) the number of reports, or sample size, n. Specifically, when n is small, one would expect ambiguity to be high; however as n increases, we would expect ambiguity to decrease. Thus, to incorporate the effects of n explicitly in our model, let $\theta = \theta'/n$ such that,

$$S(f:c) = p_A + \frac{\theta'}{n}\left(1 - p_A - p_A^{\beta}\right) \qquad (8)$$

where S(f:c) = judged probability, and $p_A = f/n$. That is, in judging the probability of an event based on f reports "for" and c "con," people are assumed to anchor on f/n, and then adjust for the unreliability of the source and the amount of data. The model in equation (8) has several implications: (1) Consider the effect of the amount of information (n) on judged likelihood. Note that $S \to p_A$ as $n \to \infty$. This means that as the amount of information increases, one becomes more certain as to the diagnosticity of the data. It is important to realize that as $n \to \infty$, S does not go to 0 or 1 as would be implied by a standard Bayesian revision model. Instead, the fact that S asymptotes at p_A parallels an analogous result in cascaded inference where, under certain symmetry assumptions, the maximum probability of a hypothesis is bounded by the reliability of the reporting source (Schum & DuCharme, 1971).

(2) Conditional on a given value of θ', the model implies that there will be trade-offs between p and n in determining judged likelihood. For example, one might find the evidence in favor of some hypothesis to be more convincing on the basis of (9:1) than (2:0). However, because S asymptotes at p_A, trade-offs of p and n will only occur at small values of n.

(3) Since $\theta = \theta'/n$, n also affects the conditions underlying the additivity of complementary probabilities. Specifically,

$$S(f:c) + S(c:f) = 1 + \frac{\theta'}{n}\left[1 - p_A^{\beta} - (1 - p_A)^{\beta}\right] \qquad (9)$$

Thus, in addition to the additivity conditions discussed in regard to equation (7), as $n \to \infty$, additivity will hold regardless of θ', β, or p_A. Of course, when n is small (meager data), adjustments will be substantial and violations of additivity will be most likely.

Experiment 1 explicitly considers the role of n in equation (8), whereas factors affecting θ' are the central concern of experiment 2.

Experiment 1

Subjects. Thirty-two subjects were recruited through an ad in the University newspaper which offered $5 an hour for participation in an experiment on judgment. The median age of the subjects was 24, their educational level was high (mean of 4.4 years of formal post-high school education), and there were 16 males and 16 females.

Stimuli. The stimuli consisted of a set of scenarios that involved a hit-and-run accident seen by varying numbers of witnesses. Moreover, of the n witnesses to the accident, f claimed that it was a green car while c claimed it was a blue car. A typical scenario was phrased as follows:

> An automobile accident occurred at a street corner in downtown Chicago. The car that caused the accident did not stop but sped away from the scene. Of the n witnesses to the

> accident, f reported that the color of the offending car was green, whereas c reported it was blue. On the basis of this evidence, how likely is it that the car was green?

Each scenario was printed on a separate page and contained a 0-100 point rating scale that was used by the subject to judge how likely the accident was caused by a particular colored car. Each stimulus contained the same basic story but varied in the total number of witnesses (n), the number saying it was a green (f) or a blue car (c), and whether one was to judge the likelihood that the majority or minority position was true. In order to sample a wide range of values of n and p, 40 combinations were chosen as follows: for p = 1, n = 2, 6, 12, 20; p = .89, n = 9, 18, 27; p = .80, n = 5, 10, 15, 20, 25; p = .75, n = 4; p = .67, n = 3, 6, 9, 12, 15, 18, 24; p = .60, n = 5, 10; p = .50, n = 2, 8, 12, 20; p = .40, n = 5, 10; p = .33, n = 6, 9, 18; p = .25, n = 4; p = .20, n = 5, 10; p = .11, n = 9, 18; p = 0, n = 2, 6, 12, 20. In addition, 8 stimuli were given twice to ascertain test-retest reliability. Thus, the total number of stimuli was 48, and they were arranged in booklet form.

Procedure

When the subjects entered the laboratory, they were told that the experiment involved making inferential judgments. Furthermore, it was stated that if they did well in the experiment (without specifying what this meant), it was likely that they would be called for further experiments. Given the relatively high hourly wage, this was thought to provide some incentive to take the task seriously. In order to avoid boredom and to reduce the transparency that judgments of complementary events were sometimes required, subjects were given 4 sets of 12 stimuli and, after completing each set, they performed a different task. All stimuli were randomly ordered within the four sets. Subjects could take as much time as they needed and they were free to

make as many (or as few) calculations as they wished. After completing the task, all subjects filled out a questionnaire regarding various demographic variables.

Estimating the Model

To estimate the model from the experimental data, we need to re-write equation (8) and include a random error term to represent judgmental inconsistency; therefore,

$$S(f:c) = p_A + \frac{\theta'}{n}\left(1 - p_A - p_A^{\beta}\right) + \varepsilon \qquad (10)$$

Equation (10) requires a non-linear estimation technique which was developed in the following way: let $S(f:c)$ be the actual response of the subject and $\hat{S}(f:c)$ the predicted response from the model. We wish to minimize some loss function (we chose the mean absolute deviation, MAD), by finding values of θ' and β such that,

$$\frac{\sum^{N}\left|S(f:c) - \hat{S}(f:c)\right|}{N} = \text{minimum} \qquad (11)$$

This was done by setting up a grid of values of θ' and β and writing a computer program to first compute the MAD for pairs of "coarse" values of θ' and β. Since certain ranges of θ' and β can thus be excluded, the program then considers "finer-grained" values until MAD is minimized.[2] The output from this analysis is a unique set of values for θ' and β that minimizes the desired loss function.

Since the sampling distributions of θ' and β are not known, testing the statistical significance of the model's fit to the data is problematic. We therefore adopted the strategy of comparing the accuracy of $\hat{S}(f:c)$ with that of a model based solely on p_A. Moreover, since p_A is the anchor of the assumed process, any difference between the accuracy of p_A and $\hat{S}(f:c)$

can be attributed to the adjustment process, and thus to θ' and β. We emphasize that this procedure is biased <u>against</u> finding differences between p_A and $\hat{S}(f:c)$ for two reasons: (a) the model predicts that $S(f:c) \rightarrow p_A$ as n increases. Thus, since we have included some large values of n to test this prediction, if $S(f:c) = p_A$, this counts against, rather than for, the model; (b) the model further predicts that $S(f:c) = p_A$ at the cross-over point, p_c, and will be close to p_A in the region of p_c. Again, if this occurs, it counts against the model. We take this highly conservative approach to guard against attributing random error in the data to an adjustment process.

Results

Before discussing the major results, recall that for each subject, 8 stimuli were given twice so that test-retest reliability could be assessed. This was done in two ways: (1) the correlation between judgments of the same stimuli, within each subject (N = 8), was computed. The mean of these correlations was .93, with 26 of the 32 coefficients greater than .90; (2) each subject was considered a replication with 8 responses and the correlation between judgments for identical stimuli, over subjects (N = 256 = 32 subjects × 8 responses), was .91. Clearly, the reliability of the judgments was high, regardless of the computational method.

For a general impression of how well the model fits the data, we first consider an aggregate analysis (individual differences will be considered in detail below). For each of the 48 stimuli, the judgments from the 32 subjects were averaged to form the mean judged likelihood, $\bar{S}(f:c)$. This was then used as the dependent variable to be fit by the model. The parameter values obtained from the estimation program were, $\hat{\theta}' = .35$, $\hat{\beta} = .10$ (implying $\hat{p}_c = .16$). In addition, the mean absolute deviation of model and data was

.020, which is significantly lower than that of the baseline p_A-model (MAD = .041; $p < .001$ using a Wilcoxon matched-pairs signed-ranks test).

To see whether the implications of the model hold, consider Table 1, which shows $\bar{S}(f:c)$ and $\hat{S}(f:c)$ for the 48 stimuli. First, does

Insert Table 1 about here

$S(f:c) \to p_A$ as n increases? The data strongly support this when $p_A = 1$, .67, .60, .50, .40, and 0. At the values of .75 and .25, n was not varied although the large adjustments do suggest that the expected effect would occur. However, the effect of n is less clear at $p = .89$, .80, and .33 since there is little initial adjustment at small n. Taken together, these results suggest moderate support for the hypothesis. Second, do p and n trade-off in affecting judged likelihood? The evidence here is quite convincing: e.g., note that $S(8:1) = .88 > S(2:0) = .85$, $S(10:5) = .65 > S(3:1) = .63$, $S(1:4) = .21 > S(1:3) = .20$. Of particular interest is the result that $S(0:2) = .16 > S(1:8) = .12$. This means that when there is limited evidence, no data in favor of a hypothesis can be judged as stronger evidence <u>for</u> that hypothesis than when more evidence is available with mixed support. Third, an important implication of the model concerns the relation between θ, β, and the additivity of complementary probabilities. Recall from equation (7) that when $\theta > 0$ and $\beta < 1$, sub-additivity is predicted for $0 < p_A < 1$. To test this prediction, consider Table 2, which shows both $\bar{S}(f:c) + \bar{S}(c:f)$ and $\hat{S}(f:c) + \hat{S}(c:f)$. Note that there is substantial sub-additivity and the model does a reasonably good job of capturing it. In

Insert Table 2 about here

judging the performance of the model in this regard, it is useful to remember that we have gone beyond the qualitative prediction that sub-additivity will

TABLE 1

Fit of the Model for Aggregate Data

n	p_A	$\bar{s}$	$\hat{s}$
2	1	.85	.83
6	1	.92	.94
12	1	.96	.97
20	1	.95	.98
9	.89	.88	.86
18	.89	.87	.87
(18)	(.89)	(.85)	.87
27	.89	.87	.88
5	.80	.80	.75
(5)	(.80)	(.73)	.75
10	.80	.79	.77
15	.80	.81	.78
20	.80	.80	.79
25	.80	.82	.79
(25)	(.80)	(.80)	.79
4	.75	.63	.69
3	.67	.61	.60
(3)	(.67)	(.59)	.60
6	.67	.62	.63
(6)	(.67)	(.63)	.63
9	.67	.61	.65
12	.67	.64	.65
15	.67	.65	.66
18	.67	.63	.66
24	.67	.66	.66
5	.60	.53	.56
10	.60	.58	.58
2	.50	.45	.42
8	.50	.44	.48
(8)	(.50)	(.47)	.48
12	.50	.47	.49
20	.50	.47	.49
5	.40	.36	.38
10	.40	.39	.39
6	.33	.31	.32
(6)	(.33)	(.29)	.32
9	.33	.27	.32
18	.33	.29	.33
4	.25	.20	.24
5	.20	.21	.20
10	.20	.19	.20
(10)	(.20)	(.18)	.20
9	.11	.12	.11
18	.11	.13	.11
2	0	.16	.17
6	0	.07	.06
12	0	.06	.03
20	0	.04	.02

Note: Numbers in parentheses are for the repeat judgments.

TABLE 2

Sub-additivity for the Aggregate Data

P_A	$(1 - P_A)$	n	Actual $\bar{S}(f:c) + \bar{S}(c:f)$	Predicted $\hat{S}(f:c) + \hat{S}(c:f)$
1	0	2	1.01	1.00
1	0	6	.99	1.00
1	0	12	1.01	1.00
1	0	20	.99	1.00
.89	.11	9	1.00	.97
.89	.11	18	1.00	.98
(.89)	(.11)	(18)	(.98)	.98
.80	.20	5	1.01	.95
(.80)	(.20)	(5)	.94	.95
.80	.20	10	.98	.97
.75	.25	4	.83	.93
.67	.33	6	.92	.95
(.67)	(.33)	(6)	(.92)	.95
.67	.33	9	.88	.97
.67	.33	18	.92	.99
.60	.40	5	.89	.94
.60	.40	10	.97	.97
.50	.50	2	.90	.84
.50	.50	8	.88	.96
(.50)	(.50)	(8)	(.94)	.96
.50	.50	12	.95	.98
.50	.50	20	.94	.98

Note: Numbers in parentheses are for the repeat judgments.

be present in the data, to specifying both the amount of the effect and the conditions under which it will not be present. Given these goals, we view the results as supporting our model. Moreover, note that the baseline p_A-model would always predict perfect additivity and thus does not describe these data well.

Individual Analyses

Since each subject rated all stimuli, we can fit the model for each person. These results are shown in Table 3. The table indicates substantial

Insert Table 3 about here

individual differences in the parameter values and the degree to which the model fits the data (as indicated by the MAD's). When compared with the aggregate analysis, the individual models contain considerably more noise (recall that the MAD for the aggregate data is .020). Furthermore, in comparing each subject's model against the baseline p_A-model, 14 of the 32 subjects showed no significant adjustment process, as specified by our model, while 18 did. The reason for the emphasis is that no subject, even those for whom $\hat{\theta}' = 0$, used a strict p_A-strategy (i.e., $S(f{:}c) = p_A$ for all p_A and n). Instead, some used p_A most of the time but occasionally adjusted for n at $p_A = 0$ and 1, while others had no clearly discernible strategy. This helps to explain why the MAD for subjects with $\hat{\theta}' < .10$ is not close to zero, as would be expected if they simply used p_A for making their judgments. Indeed, subject 6 ($\hat{\theta}' = .02$) had the highest MAD of the 32 subjects. Thus, there seem to be idiosyncratic ways of making probability judgments that are not captured by equation (8).

The above should not detract from the fact that a majority of subjects did show a significant adjustment in accord with the theory. We illustrate

TABLE 3

Fit of the Model for Individual Subjects

Ss	$\hat{\theta}'$	$\hat{\beta}$	$\hat{p}_c$	MAD	
1	.00	-	-	.051	ns
2	.00	-	-	.062	ns
3	.02	.01	.03	.002	ns
4	.02	.14	.20	.025	ns
5	.02	.20	.24	.040	ns
6	.02	.23	.27	.113	ns
7	.05	.00	.00	.007	ns
8	.10	1.00	.50	.052	ns
9	.11	.11	.17	.025	**
10	.13	.02	.06	.037	*
11	.15	.00	.00	.081	*
12	.17	.04	.09	.069	**
13	.24	.01	.03	.051	ns
14	.24	.21	.25	.031	ns
15	.28	10.90	.84	.051	***
16	.30	60.00	.95	.052	ns
17	.36	.01	.03	.052	***
18	.36	1.00	.50	.030	**
19	.37	.02	.06	.077	**
20	.37	.08	.14	.033	***
21	.37	.12	.18	.010	ns
22	.42	.04	.09	.079	ns
23	.42	.14	.20	.057	ns
24	.44	.06	.12	.027	***
25	.48	.02	.06	.088	**
26	.50	.01	.03	.023	***
27	.55	.02	.06	.046	***
28	.64	.11	.17	.053	***
29	.84	1.50	.57	.070	*
30	.93	.89	.48	.069	***
31	1.34	.01	.03	.089	***
32	1.83	.03	.08	.106	***

* $p < .05$ (Wilcoxon test)
** $p < .01$
*** $p < .001$
ns not significant

this by the results of five subjects, each representing a different combination of θ' and β parameters. This is shown in Table 4. Subject

Insert Table 4 about here

26 illustrates the use of a highly consistent strategy in which downward adjustments are made over almost the entire range of p. Subject 18 also has a consistent strategy involving adjustments, but $\hat{p}_c = .50$, implying that adjustments will be down when $p_A > .5$, up when $p_A < .5$, and no adjustments at $p_A = .5$. The data conform quite closely to this pattern. Subject 15 has a somewhat less consistent strategy of making small upward adjustments over most of the range of p ($\hat{p}_c = .84$). Again, the data are generally consistent with this interpretation. Subject 3 is included for contrast since, as can be seen, there was almost total reliance on p_A (as would be predicted by the parameter values and low MAD). Subject 32 is shown to illustrate the most extreme and least consistent adjustment process (which was generally downward). As is evident from the data, this subject had difficulty in "controlling" the adjustment process (cf. Hammond & Summers, 1972, on "cognitive control"). This lack of consistency manifested itself in widely different adjustments for the same stimuli as well as illogical judgments. An example of the latter was that evidence of (0:2) was evaluated as stronger than (2:0) (i.e., .40 vs. .30). The lack of consistency and large amount of adjusting that characterize subject 32 suggested that there might be a positive relation between the size of θ' and MAD, over subjects. When we investigated this, the correlation was $r = .46$ ($p < .001$). Thus, there seems to be a connection between the amount of adjustment and the ability to execute it consistently.

Our final results concern the additivity/non-additivity of complementary probabilities for individual subjects. This is illustrated using the subjects

TABLE 4

Fit of the Model for Selected Subjects

		Subject #26		Subject #18		Subject #15		Subject #3		Subject #32	
n	p_A	S	$\hat{S}$	S	$\hat{S}$	S	$\hat{S}$	S	$\hat{S}$	S	$\hat{S}$
2	1	.80	.75	.70	.82	.92	.86	.99	.99	.30	.09
6	1	.89	.92	.89	.94	.46	.95	.99	1.00	.80	.70
12	1	.97	.96	.90	.97	.95	.98	.99	1.00	.80	.85
20	1	.98	.98	.88	.98	.88	.99	.99	1.00	.70	.91
9	.89	.83	.94	.84	.86	.84	.89	.88	.89	.90	.71
18	.89	.84	.87	.85	.87	.87	.89	.89	.89	.80	.80
(18)	.89	.87	.87	.85	.87	.87	.89	.89	.89	.70	.80
27	.89	.85	.89	.82	.88	.82	.89	.89	.89	.60	.83
5	.80	.77	.72	.80	.76	.85	.81	.80	.80	.60	.51
(10)	.80	.72	.76	.80	.78	.83	.80	.80	.80	.90	.65
10	.80	.76	.76	.80	.78	.84	.80	.80	.80	.60	.65
15	.80	.79	.77	.79	.79	.74	.80	.80	.80	.70	.70
20	.80	.76	.78	.79	.79	.72	.80	.80	.80	.60	.73
25	.80	.76	.78	.80	.79	.85	.80	.80	.80	.90	.74
(25)	.80	.76	.78	.80	.79	.74	.80	.80	.80	.70	.74
4	.75	.57	.66	.65	.70	.72	.76	.66	.75	.30	.41
3	.67	.57	.56	.64	.63	.74	.70	.66	.67	.20	.27
(3)	.67	.61	.56	.65	.63	.75	.70	.66	.67	.20	.27
6	.67	.59	.61	.65	.65	.72	.69	.67	.67	.60	.47
(6)	.67	.57	.61	.65	.65	.72	.69	.66	.67	.30	.47
9	.67	.59	.63	.65	.66	.65	.68	.66	.67	.30	.54
12	.67	.60	.61	.64	.66	.72	.68	.66	.67	.40	.57
15	.67	.62	.65	.65	.66	.73	.68	.66	.67	.70	.59
18	.67	.63	.65	.65	.66	.65	.68	.67	.67	.70	.60
24	.67	.62	.66	.65	.66	.63	.67	.66	.67	.50	.62
5	.60	.52	.54	.60	.59	.67	.62	.60	.60	.60	.39
10	.60	.57	.57	.60	.59	.66	.61	.60	.60	.40	.49
2	.50	.38	.38	.50	.50	.57	.57	.50	.50	.20	.06
8	.50	.42	.47	.49	.50	.53	.52	.50	.50	.30	.39
(8)	.50	.50	.47	.50	.50	.52	.52	.50	.50	.30	.39
12	.50	.47	.48	.50	.50	.54	.51	.50	.50	.30	.43
20	.50	.48	.49	.51	.50	.55	.51	.50	.50	.30	.46
5	.40	.36	.36	.40	.41	.42	.43	.40	.40	.20	.26
10	.40	.40	.38	.40	.41	.34	.42	.39	.40	.50	.33
6	.33	.27	.30	.31	.35	.25	.36	.34	.33	.40	.24
(6)	.33	.30	.30	.34	.35	.38	.36	.34	.33	.20	.24
9	.33	.26	.31	.34	.34	.35	.35	.33	.33	.20	.27
18	.33	.30	.32	.34	.34	.33	.34	.33	.33	.30	.30
4	.25	.25	.22	.25	.30	.24	.30	.25	.25	.10	.15
5	.20	.18	.18	.21	.24	.23	.25	.20	.20	.20	.14
10	.20	.18	.19	.21	.22	.22	.22	.20	.20	.10	.17
(10)	.20	.20	.19	.40	.22	.26	.22	.20	.20	.20	.17
9	.11	.08	.11	.17	.14	.12	.12	.11	.11	.10	.10
18	.11	.08	.11	.15	.13	.12	.12	.12	.11	.10	.10
2	0	.05	.25	.10	.18	.14	.14	.00	.01	.40	.92
6	0	.03	.08	.10	.06	.13	.05	.00	.00	.30	.31
12	0	.02	.04	.10	.03	.14	.02	.00	.00	.20	.15
20	0	.02	.03	.11	.02	.12	.01	.00	.00	.10	.09
		$\hat{\theta}' = .50$		$\hat{\theta}' = .36$		$\hat{\theta}' = .28$		$\hat{\theta}' = .02$		$\hat{\theta}' = 1.83$	
		$\hat{\beta} = .01$		$\hat{\beta} = 1.00$		$\hat{\beta} = 10.90$		$\hat{\beta} = .01$		$\hat{\beta} = .03$	
		$\hat{p}_c = .03$		$\hat{p}_c = .50$		$\hat{p}_c = .84$		$\hat{p}_c = .03$		$\hat{p}_c = .08$	
		MAD = .023		MAD = .030		MAD = .051		MAD = .002		MAD = .106	

discussed above and whose data are displayed in Table 5. The important thing

Insert Table 5 about here

to note is that subject 26 is consistently sub-additive (and this is predicted quite well by the model); subject 18 is generally additive, as implied by $\hat{p}_c = .50$; subject 15 is super-additive, but not consistently so; subject 3 is additive; subject 32 is both highly sub-additive and inconsistent. From our perspective, these results strengthen our interpretation of the θ and β parameters, as well as the general form of the model.

A possible criticism of the above experiment is that although we investigated the responses of 32 individual subjects in depth, we only obtained responses to a single scenario. In other words, are our results simply a function of the content of the specific scenario investigated? Therefore we ran another 32 subjects using four different content scenarios but with the same numerical values as in the scenario involving the automobile accident. These scenarios involved: (1) A taste test where people had to identify a soft drink (Coke vs. Pepsi); (2) A bank robbery where witnesses said the robbers spoke to each other in a foreign language (German vs. Italian); (3) An experiment where 6 year old children had to identify words flashed on a screen (ROT vs. BED); and, (4) Experts investigating the cause of a fire (arson vs. short-circuit). Eight subjects were assigned at random to each scenario. Since the results from these four scenarios parallel those of the automobile-accident-scenario in terms of model fits (albeit with different parameter values), they are not presented here.

Experiment 2

We had two goals in conducting experiment 2. First, we wished to test systematically for the effects of source credibility and signal dissimilarity

TABLE 5

Additivity/Non-additivity of Complementary Probabilities

			Subject #26		Subject #18		Subject #15		Subject #3		Subject #32	
P_A	$(1 - P_A)$	n	$\sum$	$\hat{\sum}$	$\sum$	$\hat{\sum}$	$\sum$	$\hat{\sum}$	$\sum$	$\hat{\sum}$	$\sum$	$\hat{\sum}$
1	0	2	.85	1.00	.80	1.00	1.06	1.00	.99	1.00	.70	1.00
1	0	6	.92	1.00	.99	1.00	.59	1.00	.99	1.00	1.10	1.00
1	0	12	.99	1.00	1.00	1.00	1.09	1.00	.99	1.00	1.00	1.00
1	0	20	1.00	1.00	.99	1.00	1.00	1.00	.99	1.00	.80	1.00
.89	.11	9	.91	.95	1.01	1.00	.96	1.02	.99	1.00	.80	.81
.89	.11	18	.92	.97	1.00	1.00	.99	1.01	1.01	1.00	.90	.91
.89	.11	18	.95	.97	1.00	1.00	.96	1.01	1.01	1.00	.80	.91
.80	.20	5	.95	.90	1.01	1.00	1.08	1.05	1.00	1.00	.80	.65
.80	.80	10	.94	.95	1.01	1.00	1.06	1.03	1.00	1.00	.70	.83
.80	.20	10	.96	.95	1.20	1.00	1.10	1.03	1.00	1.00	.80	.83
.75	.25	4	.82	.88	.90	1.00	.96	1.07	.91	1.00	.40	.57
.67	.33	6	.86	.92	.99	1.00	.97	1.05	1.01	1.00	1.00	.71
.67	.33	6	.87	.92	.96	1.00	1.10	1.05	1.00	1.00	.50	.71
.67	.33	9	.85	.95	.99	1.00	1.00	1.03	.99	1.00	.50	.81
.67	.33	18	.93	.97	.99	1.00	.98	1.02	1.00	1.00	1.00	.90
.60	.40	5	.88	.90	1.00	1.00	1.09	1.06	.99	1.00	.80	.65
.60	.40	10	.97	.95	1.00	1.00	1.00	1.03	1.00	1.00	.90	.82
.50	.50	2	.76	.75	1.00	1.00	1.14	1.14	1.00	.99	.40	.12
.50	.50	8	1.00	.94	.98	1.00	1.06	1.03	1.00	1.00	.60	.78
.50	.50	8	.84	.94	1.00	1.00	1.04	1.03	1.00	1.00	.60	.78
.50	.50	12	.94	.96	1.00	1.00	1.08	1.02	1.00	1.00	.60	.85
.50	.50	20	.96	.98	1.02	1.00	1.10	1.01	1.00	1.00	.60	.91
			$\hat{\theta}' = .50$		$\hat{\theta}' = .36$		$\hat{\theta}' = .28$		$\hat{\theta}' = .02$		$\hat{\theta}' = 1.83$	
			$\hat{\beta} = .01$		$\hat{\beta} = 1.00$		$\hat{\beta} = 10.90$		$\hat{\beta} = .01$		$\hat{\beta} = .03$	

on the parameters of the model. In accordance with our theory, θ' should decrease as source credibility and signal dissimilarity increase. Second, we wished to investigate the importance of individual differences in the way people cope with the ambiguity inherent in our judgment task.

METHOD

Design. Two levels (high/low) of source credibility and dissimilarity of signals were crossed in a 2 × 2 factorial design. In addition, four different content scenarios were constructed that varied on all four experimental combinations (resulting in 16 different stories). Subjects were asked to judge 21 stimuli that varied in p and n (see below) for each of the four content-distinct scenarios. Thus, each subject initially made 84 probability judgments. However, to reduce boredom in the task, subjects made judgments in all four scenarios, with each scenario representing one of the four experimental conditions. For example, subject 1 received scenario A in the high/high condition, scenario B in the high/low condition, and so on. A four-person latin-square was set up so that every scenario appeared an equal number of times in each experimental condition. Finally, since subjects made judgments in one scenario under the high/high condition, the same scenario was also given in the low/low condition (and the order was counter-balanced). In this way, we were able to examine each subject's judgments holding the content of the scenario constant. This part of the experiment required 21 additional judgments, making the total number of responses for each subject equal to 105.

Stimuli. The four content scenarios used involved the automobile accident from experiment 1, the word-recognition task described above and two new stories. These latter scenarios involved determining the name of a play from an excerpt, and the diagnosis of a medical condition. Four versions of

each scenario were constructed to reflect different levels of credibility and dissimilarity (e.g., in the word-recognition task, we had 15 vs. 6 year olds and BED vs. ROT as opposed to BED vs. BID). Within each scenario, subjects were given 21 stimuli that reflected the amount of evidence for each hypothesis. The values of the stimuli were different from those used in experiment 1 in that smaller values of n were used in order to provide more sensitive tests of the model. The stimuli used were: for p = 0, 1, n = 1, 2, 6; for p = .125, .875, n = 8; for p = .2, .8, n = 5; for p = .25, .75, n = 4; for p = .33, .67, n = 6, 9; for p = .67, n = 3; for p = .4, .6, n = 5; for p = .5, n = 2, 8.

Subjects and Procedures. Thirty-two subjects participated in this experiment (comprising 8, 4-person latin-squares). Subjects were paid $5 per hour and the task took about one hour to complete. The tasks were presented in booklets and after each series of 21 judgments, subjects were either given a break or another task. At the end of the experiment, a manipulation check was performed on the credibility and dissimilarity induction. Specifically, each subject was asked to rate (using a 0-100 scale) the credibility of the source and the confusability of the signals in all four scenarios. Since each scenario had high and low levels of each factor, the subjects rated credibility and dissimilarity under both conditions. Therefore, subjects made 4 judgments on each of the 4 scenarios.

Results

Before presenting the main results, we note that the manipulation check showed that subjects did, on average, see the "high" credibility versions of the same scenarios as greater than the low (80 vs. 47); and the high dissimilarity signals as less confusable than the low (30 vs. 62).

(1) General fit of the model: For each subject in each experimental condition, the model was fit to yield estimates of θ' and β (this resulted in 160 models - 32 subjects × 5 models). The fit of the individual models was comparable to that of experiment 1 (median MAD = .042 over all conditions).

(2) Manipulation of θ': The appropriate analysis-of-variance (2 × 2 × latin-square) was performed using $\hat{\theta}'$ as the dependent variable and the results showed a significant main effect for "credibility" ($p < .001$), no main effect for "dissimilarity," and a three-way interaction of scenario × credibility × dissimilarity ($p < .02$). The results for the main effect are shown in Table 6. The table shows that θ' does increase as the credibility

Insert Table 6 about here

of the source decreases, thereby confirming our prediction. However, there was no effect for dissimilarity, contrary to our prediction. The three-way interaction showed that in two scenarios, the effect of dissimilarity of the signals had a large effect on θ' when credibility was low, while in the other two scenarios, dissimilarity had a large effect when credibility was high. However, it is not clear why this occurred and we do not consider it further.

In addition to the above analysis, recall that each subject also received the same scenario in the high/high and low/low conditions. A comparison of the means of the estimated θ''s in these two conditions also showed a significant difference in the hypothesized direction; i.e., $\bar{\theta}' = .17$ in the high/high condition, $\bar{\theta}' = .29$ in the low/low ($p < .004$ by a paired t-test). Thus, with the exception of an effect for the dissimilarity of the signals, our hypotheses concerning θ' are supported by the experimental data.

(3) Individual differences: We now consider the following: (a) can subjects be characterized as having a general strategy, as measured by the

TABLE 6

Experiment 3 - Mean $\hat{\theta}'$ Parameters
by Experimental Conditions

Credibility	Dissimilarity: High	Dissimilarity: Low	
High	.17	.20	.19
Low	.31	.29	.30
	.24	.25	

consistency of their θ' and β values, in different scenarios?; (b) is the amount of one's adjustment, as measured by θ', systematically related to the consistency of executing one's strategy?; (c) can individual perceptions of the credibility of the source and the dissimilarity of the signals account for variance in θ' and β within each of the experimental conditions?

(a) Recall that for each subject, four different scenarios were given and a model fit to the data in each. Therefore, each subject can be characterized by four θ''s, β's, and MAD's. To determine if the parameter values were more alike within a subject than between subjects (this is measured by the intra-class correlation), a one-way repeated analysis-of-variance was performed (32 × 4) for $\hat{\theta}'$, $\hat{p}_c$, and MAD (Winer, 1963, chap. 3). The results showed that for $\hat{\theta}'$, $r = .73$ ($p < .001$); for $\hat{p}_c$, $r = .68$ ($p < .001$); and for MAD, $r = .86$ ($p < .001$). These results are particularly striking when one recalls that the four scenarios varied over the four experimental conditions. However, in spite of these differences, the results show strong and stable individual strategies in the amount that is adjusted (θ'), the direction of the adjustments (p_c or β), and the consistency of executing one's strategy (MAD).

(b) In experiment 1, we found a significant positive correlation between $\hat{\theta}'$ and MAD. The same positive relation was found here in three of the four scenarios ($r = .67, .48, .40, .10$). Thus, our interpretation of θ as reflecting a cognitive simulation process is strengthened by the generality of this finding.

(c) Since each subject made independent judgments of the credibility and confusability of the experimental stimuli, we were also able to investigate how these judgments related to θ' within experimental conditions. To do so, we re-analyzed our data with a regression model where θ' was the dependent

variable, and the individual ratings of credibility and confusability, together with dummy variables representing the different scenarios, were the independent variables. More precisely, there is a regression equation of this type for each of the four experimental conditions. However, these four equations can be estimated more efficiently as a single model using Zellner's (1962) procedure for "seemingly unrelated" regressions. The multiple R estimated by this procedure was .44 (with an adjusted R of .35). Of the independent variables, there was no effect for either scenarios or confusability. However, all four coefficients for credibility in the different experimental conditions were significant ($p < .02$) and of the hypothesized sign (i.e., a negative relation between θ' and ratings of credibility). We interpret these results as strengthening the conclusions drawn from the more standard ANOVA of our study; that is, θ' is not only affected by different levels of credibility across all subjects, it also covaries significantly with individual perceptions of credibility within each of these levels.

Experiment 3

The purpose of this experiment was to answer the following question: Can individuals' choices between gambles be predicted from knowledge of their θ' and β parameters obtained from a separate inference task? To examine this, subjects were first asked to make judgments as in experiments 1-2 and both θ' and β were estimated as before. The subjects were then asked to choose (or express indifference) between 9 pairs of gambles involving the outcome from an urn with known probability versus the occurrence of an event on the basis of unreliable reports. If θ' and β do capture aspects of ambiguity that affect choice, knowledge of these parameters should allow one to predict individual choices in addition to inferences.

Subjects. Twenty subjects, recruited from the University of Chicago community, participated in this study. They were paid $5/hour.

Stimuli. For the inference task, two different scenarios were used: the automobile-accident story, and the taste-test story (Pepsi vs. Coke) for which we had also previously collected data (see end of experiment 1). These were chosen because the θ' and β values were quite different in the two cases. In both scenarios, subjects received 40 combinations of p and n that were identical to those used in experiment 1. The stimuli for the choice task involved one of the following: (a) In the automobile-accident task, subjects were faced with choosing between betting that a ball drawn from an urn with known probability was green, versus, betting that the car that caused the accident was green based on witnesses' reports of the car color; (b) For those in the taste-test scenario, the choice was similarly between betting that the outcome from an urn was a certain color, versus, betting that the drink was Pepsi-Cola. In both scenarios, subjects were told to imagine that their payoff for being correct would be $10. Thus, the payoffs for the urn gamble and the bet involving the report of some event were equal. Within scenarios, each subject saw 9 pairs of gambles that varied in the proportion of colored balls in the urn and the proportion of reports favoring the particular hypothesis. These proportions were always the same in the two bets. The exact values of p used in the 9 pairs were: 1, .875, .75, .625, .50, .375, .25, .125, and 0. The number of balls in the urn and the number of reports were held constant at 8.

Procedure. The 20 subjects were randomly assigned to one of the two scenarios. The procedure for the inference task was identical to the previous experiments. After subjects finished the inference task, they were presented with the appropriate choice task. The nature of the two gambles was

explained, and subjects were then asked to choose, or indicate indifference, between the gambles. If they were not indifferent, they were also asked to indicate their strength of preference on a 4-point scale (from "little" to "great deal"). After doing this for one value of p, they turned the page and made another choice (and strength of preference rating, if appropriate) at the next level of p. This continued until all 9 pairs had been considered. Therefore, for each subject, there were 9 choices between an unambiguous bet from an urn with known p, versus an ambiguous bet that an event occurred, on the basis of the proportion of favorable reports from an unreliable source.

Results. Since each subject first participated in the inference task, we briefly consider these results before discussing the choice data. As expected, there were marked differences in the θ' and β parameters in the two scenarios. The medians for θ' and p_c (implied by β) were .13 and .11, respectively, in the automobile-accident scenario. For the taste-test story, the median θ' was 1.35 and median $p_c = .45$. Thus, the taste-test scenario induced much adjustment, with a cross-over point near .50, while the automobile-accident story induced less adjustment but a lower cross-over point.

To compare each subject's choices with predictions from the inference model, the following procedure was used: any combination of θ' and p_c implies when and where $S(p_A)$ is greater, less than, or equal to, p_A (see equation (8)). Thus, for each subject, when $p_A > S(p_A)$, we predicted the urn would be chosen over the bet based on unreliable reports; when $S(p_A) > p_A$, the opposite prediction was made; when $S(p_A) = p_A$, we predicted indifference between the two gambles. Note that when $\theta' = 0$, we always predicted indifference between the gambles since $S(p_A) = p_A$ for all p_A. In Table 7, we show the $\hat{\theta}'$ and $\hat{p}_c$ values for each subject (grouped by

Insert Table 7 about here

scenario), and the number of correct choice predictions by subject.

To evaluate how well the choices were predicted from knowledge of $\hat{\theta}'$ and $\hat{p}_c$, we used a random baseline for comparison; i.e., for each of the 9 choices made by a subject, there were three possible outcomes; urn, report, or indifference. Since the probability of randomly predicting the correct response is 1/3, we computed the probability of getting at least r hits in 9 trials on the basis of chance (using the binomial distribution). This probability is shown in the last column of Table 7. For example, subject 1 was correctly predicted in 8 of the 9 choices; the probability of getting at least this many hits by chance is .001. Thus, we rejected the hypothesis that our predictions for this subject were no better than chance. Using this method for all subjects, it can be seen that 5 of the 10 subjects in the automobile-accident scenario, and 4 of 10 in the taste-test, are well predicted using a type I error level of .05. If this error level were increased to .15, a majority of subjects (12/20) would be accurately predicted from their inference parameters. In any event, at the aggregate level (over subjects and scenarios), there were 103 hits out of 179 predictions (one response was missing). The probability of getting at least this many hits by chance is miniscule.

Second, consider the results concerning the strength of preference ratings. Recall that in addition to choosing between gambles, subjects were asked to rate their strength of preference on a 4-point scale. These ratings supplement our analysis of the choice data in the following way: in each scenario, the number of prediction errors was 38. However, in the taste-test, θ' is much larger than in the automobile-accident scenario. Since θ' is directly related to the amount of adjustment to p_A, the differences

TABLE 7

Choice Predictions from Knowledge of $\hat{\theta}'$ and $\hat{p}_c$

	Ss	$\hat{\theta}'$	$\hat{p}_c$	No. of hits	Prob. (r > hits)
	1	.00	-	8	.001
	2	.00	-	1	.849
	3	.00	-	3	.612
	4	.00	-	4	.341
Automobile-Accident	5	.05	.07	7	.008
Scenario	6	.10	.74	5	.140
	7	.16	.15	6	.040
	8	.19	.20	6	.040
	9	.63	.02	5	.140
	10	.75	.50	6	.040
	11	.20	.21	7	.008
	12	.24	.07	4	.341
	13	.66	.50	4	.341
	14	.80	.50	1	.970
Taste-Test	15	.96	.50	5	.140
Scenario	16	1.50	.50	7	.008
	17	1.71	.40	7	.008
	18	1.99	.50	5	.140
	19	2.01	.06	9	.000
	20	3.20	.02	3	.612

between $S(p_A)$ and p_A should be larger in the taste-test than in the accident story. Furthermore, the larger the differences, the stronger one's preferences should be since they are further away from indifference (where $p_A = S(p_A)$). We tested this by comparing the mean strength-of-preference ratings in the two stories across the nine levels of p. These results are shown in Table 8. First, note that the means for the taste-test are larger

Insert Table 8 about here

than the automobile-accident at every level of p. Second, the pattern of means is consistent with the general form of the model in that preferences are strongest at $p = 1$, decrease as p approaches p_C, and then increase again at $p = 0$. Therefore, the strength-of-preference data are consistent with both the difference in the sizes of θ' for the two scenarios, as well as the general form of the model.

As the astute reader may have noticed, our theory does not necessarily imply exact equivalence between choice and inference tasks since these could differ with respect to the β parameter. In particular, while payoffs are explicit in the choice task (i.e., a gain of \$10), there are no explicit payoffs in the inference task. Thus, one might expect a systematic bias between β as estimated in the inference task, and β as implied by subjects' choices. Specifically, as stated after first presenting our model, if the effect of ambiguity is to induce caution rather than riskiness, then the prospect of a gain would focus attention more on smaller rather than larger values of p(Gain) such that $\beta_{choice} < \beta_{inference}$. (Conversely, the prospect of a loss would imply more attention being paid to greater rather smaller values of p(Loss) such that $\beta_{choice} > \beta_{inference}$). Consequently, one would expect ambiguity avoidance over a wider range of p in tasks involving choice as opposed to inference. Indeed, some of the errors in

TABLE 8

Means of Strength-of-Preference
for Two Scenarios

p	Automobile Accident	Taste-test	Both Scenarios
1	3.1	3.7	3.40
.875	2.5	3.1	2.80
.750	2.0	2.5	2.25
.625	1.6	2.1	1.85
.500	.9	1.7	1.30
.375	1.4	2.1	1.75
.250	1.3	2.1	1.70
.125	.9	2.0	1.45
0	1.8	2.2	2.00
	1.72	2.39	2.05

predicting subjects' choices can be attributed to precisely this source of systematic bias. Consider the data for the automobile-accident scenario in Table 7. Eight of the 10 subjects had low $\hat{p}_c$ values in the inference task and thus could be thought of as having already conceptualized the task in terms of "gains." On the other hand, for the two subjects with high $\hat{p}_c$'s in the inference task (numbers 6 and 10), all 7 prediction errors out of 18 choices were in the same direction, namely subjects' $\hat{\beta}$'s estimated in the inference task indicated larger $\hat{p}_c$'s than were revealed by their choices. The same bias was also found in the taste-test scenario. That is, consider again only those subjects with high $\hat{p}_c$'s (numbers 13 through 18). For 4 of these subjects, all 15 out of 15 prediction errors are consistent with the $\beta_{choice} < \beta_{inference}$ bias. The two prediction errors of one subject (number 16) are in the opposite direction, and the 8 prediction errors of subject 14 are equally distributed in both directions. To summarize, we conclude that whereas individuals' parameters in an inference task can be used to predict choices, many errors of prediction are in accord with a systematic bias in the β parameter that is consistent with our theory.

Experiment 4

Having manipulated the θ parameter in experiment 2, the purpose of experiment 4 was to investigate the effects of manipulating β. This was done by allocating subjects to different roles (sellers and buyers) in an insurance context. The dependent variable of interest involved statements of maximum buying prices and minimum selling prices. The data were collected as part of a larger investigation by Hogarth and Kunreuther (1984) on the effects of ambiguity in insurance decision making.

The assumption underlying the experimental manipulation is that a person who assumes a risk, is likely to pay more attention to larger values of

p(Loss) than someone who transfers the risk. Experimental evidence consistent with this assertion has been documented by Hershey, Kunreuther, and Schoemaker (1982) and Thaler (1980). In our framework, it implies that $\beta_{seller} > \beta_{buyer}$. Given this assumption, approximate ambiguity functions for buyers and sellers of insurance can be sketched as in Figure 4. Note that when buyers/sellers in

Insert Figure 4 about here

a non-ambiguous situation, $S(p_A) = p_A$ and all responses are on the diagonal.

If one further assumes that buying and selling prices are monotonically related to $S(p_A)$, Figure 4 suggests the following predictions: (1) When buyers and sellers are in a non-ambiguous situation, $S(p_A) = p_A$, and the seller's price should equal the buyer's; (2) When buyers and sellers are equally ambiguous (i.e., their θ's are equal), the seller's price should exceed the buyer's over the whole range of p_A. Note that this arises because the seller always weights imaginary values of p(Loss) larger than the initial estimate more than the buyer. (3) Consider a seller who has no ambiguity about the probability of a loss, but a buyer who does. In Figure 4, this is shown by comparing the buyer-ambiguous function with the diagonal (seller-unambiguous). Note that the buyer's function is above the diagonal for $p_A < p_C$. This means that the buyer will perceive the probability of loss as higher than the seller, and should be willing to pay more than the seller would ask. However, when $p_A > p_C$, the buyer will perceive the loss probability as lower than the seller, and offer less than the seller would ask. This implication of the model provides a particularly stringent test for our theory. Experiment 4 was designed to test the above three predictions.

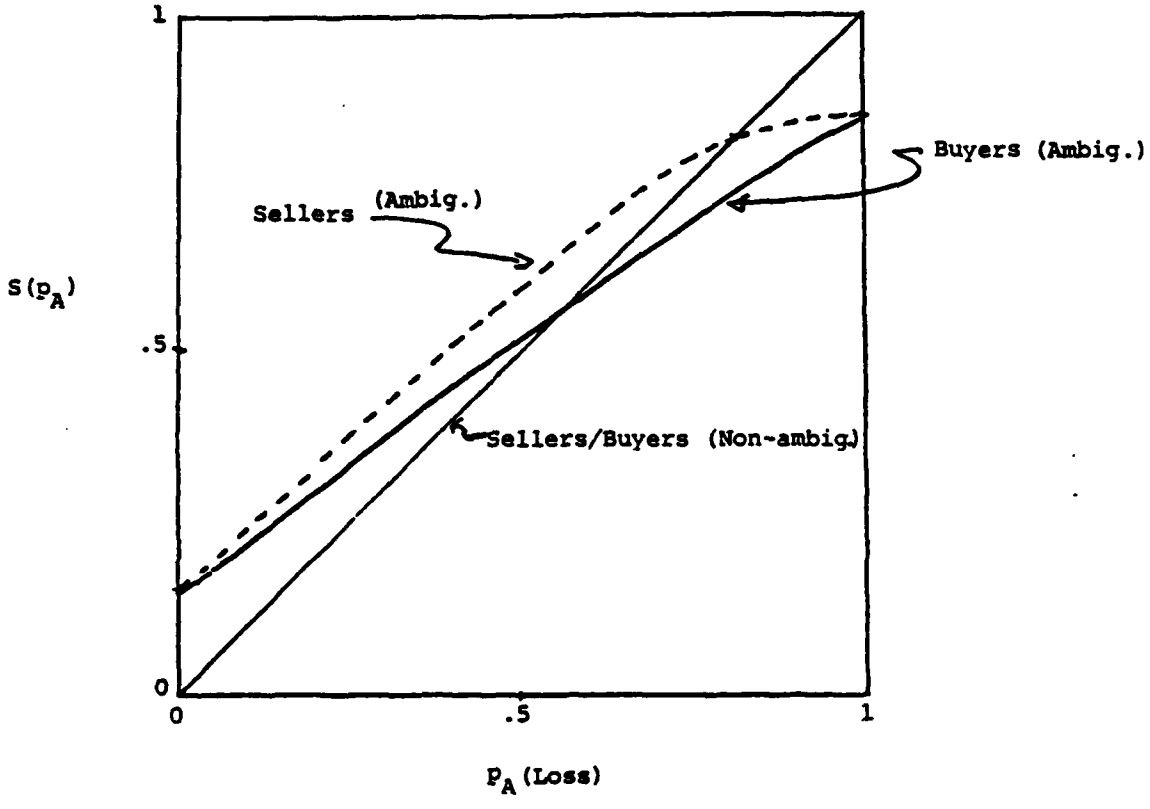


Figure 4. Approximate ambiguity functions for buyers and sellers of insurance.

METHOD

Design. Prices for insurance contingent on ambiguous and non-ambiguous probabilities were investigated across four different probability levels (.01, .35, .65, and .90). Each subject was assigned the role of buyer or seller of a contract concerning a potential $100,000 loss and responded to both ambiguous and non-ambiguous versions of the stimulus at one probability level. Thus, the design of the experiment involved three factors, two of which were between subjects (i.e., role of buyer or seller, and probability level) and one within subjects (i.e., ambiguous vs. non-ambiguous probabilities).

Stimuli. The scenario used in the stimulus material involved the owner of a small business (net assets of $110,000) who was seeking to insure against a $100,000 loss that could result from claims concerning a defective product. Subjects assigned the role of buyers were told to imagine that they were the owner of the business. Subjects assigned the role of sellers were asked to imagine that they headed a department in a large insurance company and were authorized to set premiums for the level of risk involved. Ambiguity was manipulated by factors involving how well the manufacturing process was understood, whether the reliabilities of machines used in the process were known, and the extent to which manufacturing records were well kept. In both ambiguous and non-ambiguous cases a specific probability level was stated (e.g., .01); however a comment was also added as to whether one could "feel confident" (non-ambiguous case) or "experience considerable uncertainty" (ambiguous case) concerning the estimate. As far as possible, the same wording was used in both the buyer and seller versions so that perceptions of ambiguity would be uniform in the two cases.

Subjects and procedures. Subjects were 111 MBA students at the University of Chicago who responded to questionnaires distributed in a course

on decision making. To avoid prior influence, the experiment took place during the beginning of classes. Subjects were asked to respond to the questionnaire in anonymous fashion and promised group-level feedback at a later class session (which they subsequently received). It is important to note that subjects had prior training in business, economics, and statistics, and the insurance context was familiar to them. Eight different forms of the stimulus materials, corresponding to the 2 (roles) × 4 (probability levels), were shuffled and distributed in the classrooms thereby ensuring random allocation of subjects to conditions. After reading each stimulus, subjects were asked to state maximum buying prices (for buyers) or minimum selling prices (for sellers).

Results. Table 9 reports medians for all experimental conditions as well as the differences between the sellers and buyers for the ambiguous and non-

Insert Table 9 about here

ambiguous cases respectively. We report medians since several distributions within conditions are quite skewed, and variances also differ between cells at the same probability levels, often significantly. The pattern of results in Table 9 supports our three predictions. First, in comparing buyers and sellers in the non-ambiguous case, note that the median prices are quite similar over the four probability levels. Second, when buyers and sellers are both ambiguous, observe that the selling price is considerably larger than the buying price at every level of p. This result strongly confirms the notion that $\beta_{seller} > \beta_{buyer}$ when considering ambiguous loss probabilities. Third, consider the ambiguous buyer and the non-ambiguous seller. As expected, when p is small (.01), the ambiguous buyer is willing to pay more ($1,500) than the non-ambiguous seller asks ($1,000). However, as the probability of loss increases, the two prices converge (at p = .35), and then diverge, with

TABLE 9

Medians of Selling and Buying Prices ($)

Probability levels		Sellers[1] Ambiguous		Sellers[1] Non-ambiguous	Buyers[1] Ambiguous		Buyers[1] Non-ambiguous	Difference Sellers - Buyers[2] Ambiguous	Difference Sellers - Buyers[2] Non-ambiguous
.01	Median	2500	**	1000	1500	*	1000	1000ns	0ns
	(n)		(14)			(13)			
.35	Median	52500	**	37500	35000	ns	35000	17500 **	2500ns
	(n)		(14)			(15)			
.65	Median	70000	ns	65000	45000	**	65000	25000***	0**
	(n)		(13)			(14)			
.90	Median	90000	ns	90000	60000	**	82500	30000 **	7500**
	(n)		(14)			(14)			

Two-tailed tests used throughout:

[1]Wilcoxon test.
[2]Mann-Whitney test.

* p < .05
** p < .01
*** p < .001
ns not significant

the buyer's price being less than the seller's (at p = .65, .90). Indeed, at higher probabilities, the buyer is willing to spend considerably less than the seller wants to charge. Therefore, although the buyers seem to be ambiguity avoiding at low probabilities, they paradoxically appear to be ambiguity seeking for high loss probabilities. Both results are predicted by our model. In Hogarth and Kunreuther (1984), the results of several other related experiments are reported using different scenarios, research designs, subjects, and response modes. The results of these experiments are consistent with those reported here, thus attesting to the stability of the phenomena.

DISCUSSION

We now discuss the implications of our theory and results with respect to the following issues: (1) the importance of ambiguity in assessing uncertainty; (2) the use of cognitive strategies in probabilistic judgments under ambiguity; (3) the role of ambiguity in risky choice; and, (4) extensions of the model to multiple sources and time periods.

Ambiguity and the Assessment of Uncertainty

The concept of ambiguity highlights the distinction between one's lack of knowledge of the process that generates outcomes and the uncertainty of outcomes conditional on some model of the process. The fact that there are at least two sources of uncertainty in most situations leads to the irony that one needs a well-defined model to give precise estimates of how much one doesn't know. Indeed, the usefulness of formulating well-defined stochastic processes is in eliminating ambiguity so that outcome uncertainty can be quantified. Thus, when coins are "fair" or random drawings are taken from urns with known p, there is no second-order uncertainty. Furthermore, the conditional nature of uncertainty is implicitly recognized in various attempts

to quantify and improve inferential judgments. For example, consider how uncertainty is defined in the "lens model" (Hammond, et al., 1964). In this case, the uncertainty in the environment is measured as the residual variance not accounted for by a well-formulated ecological model. Thus, unexplained variance or uncertainty is conditional on the model of how particular cues combine to form the criterion of interest. Now consider the work of Nisbett and colleagues on trying to improve probabilistic judgments through training (Nisbett, et al., 1983; Jepson, et al., 1983). They argue that training and experience can allow one to see the underlying structure of real-world problems so that the appropriate model can be used for making better judgments. Thus, the focus of their training is on making various statistical principles (e.g., regression-to-the-mean, law of large numbers, use of base rates, etc.) more obvious in everyday inferences.

While the conditional nature of uncertainty has been implicitly recognized, ambiguity results from its explicit recognition; i.e., by realizing that the "model" is itself subject to uncertainty. Indeed, one could argue that the cost of urn models, coin-flipping analogies, and the like, is that they obscure the fact that most real world generating processes are not precisely known. Furthermore, even if a process is well-defined at one point in time, the parameter(s) of the process can change over time, resulting in ambiguity as well as uncertainty. For example, imagine that you have been asked to evaluate the research output of a younger colleague being considered for promotion. Your colleague has produced 11 papers; of these the first 9 (in chronological order) represent competent, albeit unexciting scholarly work. On the other hand, the latter 2 papers are quite different; they are innovative and suggest a creativity and depth of thought absent from the earlier work. What should you do? As someone who is aware of regression

fallacies, you might consider the two outstanding papers as outliers from a stable generating process and thus predict regression-to-the-mean. Alternatively, you might consider the outstanding papers as "extreme" responses that signal a change in the generating process; i.e., a new and higher mean. If this were the case, the same general regression model would predict future papers of high quality (regression to a higher mean). If one asks what is the nature of the signaling in this case, it is obvious that the chronological order of the papers is crucial. Indeed, imagine that the outstanding papers were the first two that were written; or consider that they were the second and sixth. Each of these cases suggests a different underlying model and perhaps a different prediction. In any event, the uncertainty associated with any prediction is complicated by the ambiguity regarding the appropriate mean of the regression process.

Cognitive Strategies in Inferences Under Ambiguity

We have assumed that people use an anchoring-and-adjustment strategy in making inferences under ambiguity. However, whereas the term, "anchoring-and-adjustment" is quite general and could encompass a wide range of models (cf. Lopes, 1981; Einhorn & Hogarth, 1984), we have been quite specific as to the nature of this process in our tasks. Of greatest interest in this regard is the idea that adjustments are based on a mental simulation in which "what might be," or, "what might have been," is combined with "what is" (the anchor). The rationale for this comes from the fact that the evaluation of evidence often involves an implicit comparison process (similar to the perception of figure against ground). Thus, when evaluating the strength of evidence for a particular hypothesis, the evidence that might have been can serve as a convenient contrast case for comparison. Furthermore, since ambiguity implies that multiple models could have produced the observed

results, it seems natural to consider that different results could have occurred on the basis of different underlying processes.

The support for the hypothesized anchoring-and-adjustment strategy comes from several sources. First, recall that in our model, the largest adjustments to the anchor occur when evidence is meager. Moreover, as n increases, $S(f:c)$ asymptotes at p_A. The results of experiments 1 and 2 support this prediction. Thus, the weight of evidence (to use Keynes' term) for "what is," dominates "what might have been" as the absolute amount of evidence increases. Furthermore, the effect of increasing n is to reduce the amount of non-additivity of complementary strengths. Since most of our subjects were sub-additive, our model provides a psychological link to concerns expressed by others regarding the appropriateness of additivity when evidence is meager (Shafer, 1976; Cohen, 1977). In particular, Cohen (1977, chap. 3) points out that when one considers an incomplete system, the lower benchmark on provability is not necessarily disprovability, but nonprovability. For example, if one has meager circumstantial evidence such that the probability of the truth of a particular theory is .2, does this imply that the theory is false with $p = .8$? Rather, one might say that the theory is not proven (in a probabilistic sense) as opposed to saying that there is a .80 chance that it is wrong. Furthermore, the idea that the complement of statements can lead to "not-proved" rather than "disproved," seems to be deeply imbedded in the Anglo-American legal system. Indeed, in Scottish law, defendants are either found guilty, not-guilty, or "not proven." The last category is reserved for those cases where the evidence is too meager to allow for a judgment of guilt or innocence.

Second, the fact that non-additivity results from a shift in the direction of the adjustment process is consistent with other "order effects"

due to the use of anchoring-and-adjustment strategies. For example, in a traditional Bayesian revision task, Lopes (1981) found that a change in the order in which sample information was presented affected overall judgments by changing the anchor. Thus, consider having to judge whether samples come from an urn containing predominantly red or blue balls (70/30 in both cases). You first draw a sample of 8 that shows (5R, 3B). Thereafter, you draw another sample of 8 with the result (7R:1B). After each sample, you are asked how likely it is that you have drawn from the predominantly red urn. When the sample evidence is in the order given here, people seem to anchor on the first sample (5:3) and then adjust up for the second (stronger) sample. However, when the order of the samples is reversed, people anchor on (7:1) and adjust down for the weaker, second sample. This effect cannot be accounted for by assuming that people are using a Bayesian procedure (which treats the two situations as equal), but it does follow from an anchoring and adjustment process in which the anchor is weighted more heavily than the adjustment.

Third, the results of experiment 2 provide important evidence regarding the process assumed to underlie the model. In addition to the fact that the experimental manipulation of source credibility affected θ' as predicted, two other results were found; a positive correlation between θ' and MAD and, the stability of individual differences in θ', β, and MAD across scenarios. The first result bears directly on the nature of the adjustment process since it suggests a "cost" of engaging in mental simulation; namely, a concomitant lack of control over one's strategy (Hammond & Summers, 1972). The second result suggests strong personal propensities in evaluating evidence that transcend the particular content of scenarios. While it is too early to explicate the nature of these individual differences, their existence lends support to the idea that the parameters of our model do capture important

aspects of the process that determines judgments under ambiguity.

While our model accounts for the rather simple inferences we have studied, it also relates to an important class of inferences that result from "surprise." Consider Figure 5, which shows one's expectations for p as a

Insert Figure 5 about here

function of the credibility of the source and the dissimilarity of the signals. First, note that when credibility and dissimilarity are high, one expects p to be very high or low (recall our earlier example of cameras taking pictures of a bank robber). However, imagine that one camera showed the bank robber to be white, and the other showed him to be black. Such a result, where $p = .5$, would be surprising given the credibility of cameras and the dissimilarity of white and black robbers. Indeed, the data "are not good enough," which is represented by the range of p indicated by the two-headed arrow. Second, consider the low credibility-low dissimilarity situation; e.g., the taste-test scenario. Imagine that you were told that of the 20 people in the Pepsi vs. Coke taste-test, all correctly identified the drink as Pepsi. Such a result, where $p = 1$, would be surprising. However, this type of surprise is one where the "data are too good" rather than not good enough. Thus, there are two types of surprise and both occur when ambiguity is low. Indeed, when ambiguity is high, expectations are weak and surprise (which results from a violation of expectations) is unlikely. This situation characterizes the off-diagonal cells in the figure and accounts for our labeling these "little surprise."

Although our conceptual scheme makes clear when surprise is likely to occur, it cannot handle the variety of possible reactions it can engender. For example, when data are not good enough, it is possible to reduce the credibility of the source (e.g., the cameras were broken), synthesize the hypotheses (there were two bank robbers, one white and the other black), or

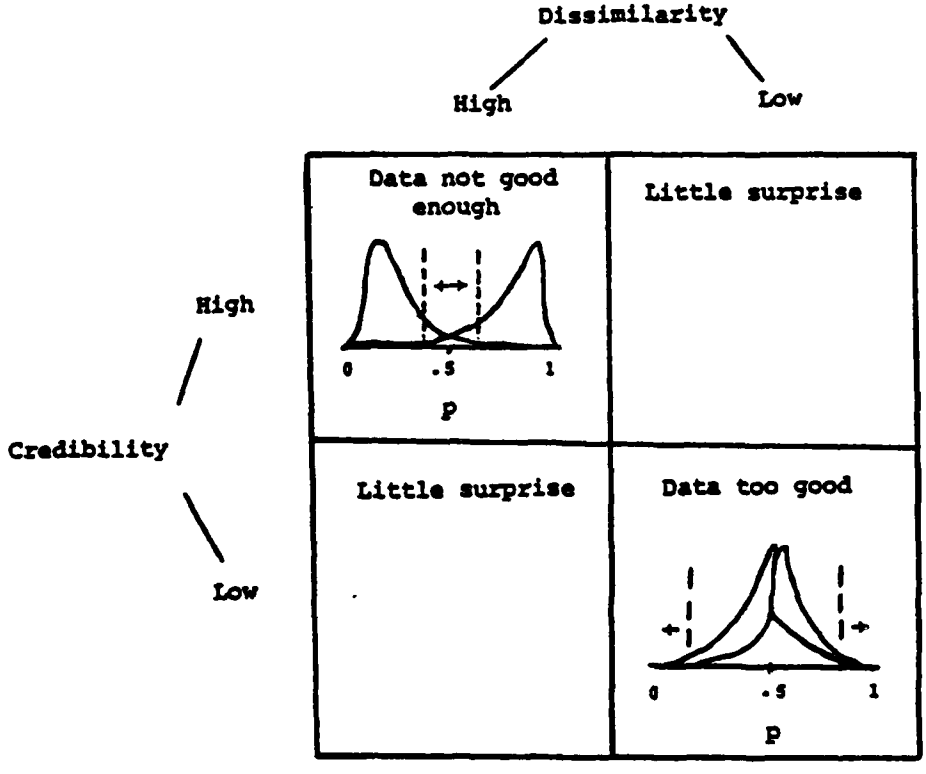


Figure 5. Expectations and surprises

otherwise make sense of the data by changing the story (e.g., there were two bank robberies on successive days). On the other hand, when data are too good, inferences of fraud, collusion, and the like, are possible (see, e.g., Kamin, 1974 on Burt's twin data; Bishop, Fienberg, & Holland, 1975, on Mendel's pea experiments). An interesting aspect of such inferences is that the surface meaning of the data can suggest the opposite conclusion; e.g., consider someone who "protesteth too much," or a suspect who was "framed" for a crime. Indeed, this is implied by our model. Specifically, consider the case of totally unreliable data which imply $\Theta = 1$ (see equation (6)). In this case,

$$S(p_A) = 1 - p_A^{\beta} \qquad (11)$$

Thus, as p_A increases, $S(p_A)$ decreases. More generally, as Θ increases, it will reach a point, conditional on p_A and β, where the evidence for a hypothesis will start to be counted against it.

Ambiguity and Risk

Although the importance of ambiguity for understanding risk has been evident since Ellsberg's original article, its omission from the voluminous literature on risk is puzzling. One reason may be the reliance on the explicit lottery, with stated payoffs and probabilities, for representing risky choice. Indeed, as Lopes (1983) has noted,

> The simple, static lottery or gamble is as indispensable to research on risk as is the fruitfly to genetics. The reason is obvious; lotteries, like fruitflies, provide a simplified laboratory model of the real world, one that displays its essential characteristics while allowing for the manipulation and control of important experimental variables. (1983, p. 137)

It should be further noted that the explicit lottery has been of equal importance to those interested in axiom systems and formal models of risk.

While explicit lotteries have been, and continue to be, useful for

studying risk, the ambiguities surrounding real world processes in domains such as nuclear power, environmental safety, and the like, accentuate the incomplete nature of such representations. Indeed, Ellsberg pointed out the particular importance of ambiguity in understanding people's reactions to new technologies (also see, Edwards & von Winterfeldt, 1982, for a historical look at reactions to earlier technological innovations). In any event, the neglect of ambiguity in theories of risk is slowly giving way to interest at both the formal-axiomatic level (e.g., Fishburn, 1983a, 1983b; Gärdenfors & Sahlin, 1982; 1983; Morris, 1983) as well as the psychological level (Lopes, 1983).

Our model of inference under ambiguity has several implications for descriptive models of risky choice. First, since the β parameter can be related to the desirability of outcomes, the model implies a form of utility × probability interaction. Moreover, experiment 4 provides direct evidence for this interaction. However, the utility × probability interaction only has an effect in the presence of ambiguity, i.e., when $\theta > 0$. Thus, whereas the bilinear assumption may be appropriate for models that exclude the effects of ambiguity (e.g., Kahneman & Tversky, 1979), it is not clear that this assumption can be maintained when ambiguity prevails. Second, both our model and data show that the net effect of the adjustment process (i.e., k) varies in magnitude with the level of p_A. Thus, theories of inference that weight probabilities according to some "reliability" factor (e.g., Gärdenfors & Sahlin, 1982) need to consider this interaction explicitly to achieve descriptive realism. Third, the model highlights the difficulty of inferring underlying attitudes toward risk from choices made in ambiguous circumstances. For example, a person buying insurance against a potential loss that is contingent on a small, ambiguous probability might appear risk averse; however, the same person could appear to be risk-seeking if the probability

were larger (cf. experiment 4). Viewed from the framework of expected utility theory, such behavior would imply an inconsistent utility function. However, this need not be the case since the apparent changes in risk attitude could result from the effects of ambiguity. At the very least, our model provides a way of analyzing the sources of such seemingly inconsistent behavior. As Hogarth and Kunreuther (1984) point out, scholars have often attempted to resolve anomalous choice patterns by considering different forms of utility functions. On the other hand, transformations of probabilities have received far less formal attention (for an exception, see Karmarkar, 1978). Finally, whereas our model does not explicate all aspects of ambiguous choice, it does suggest exciting possibilities for further work in this area.

Extensions to Multiple Sources and Time Periods

To examine inferences under ambiguity in depth, we have restricted ourselves to how evidence from a single source is evaluated at one point in time. However, consider the more realistic situation where decision makers receive information from multiple source-types (including base rates) over multiple time periods. The aggregation of information over source-types and time can be conceptualized by an "evidence matrix" that has source-types for rows and time periods for columns. Such a matrix is shown in Figure 6. The

Insert Figure 6 about here

entries in each cell of the matrix reflect the conflicting evidence received from a source-type in that period. The matrix provides a simple yet powerful way to look at a wide variety of inference problems. In particular, by focusing on source-types (rows) or time periods (columns), one can look at the combining of information either longitudinally, cross-sectionally, or both. Furthermore, the issues of reliability and ambiguity become quite complex here

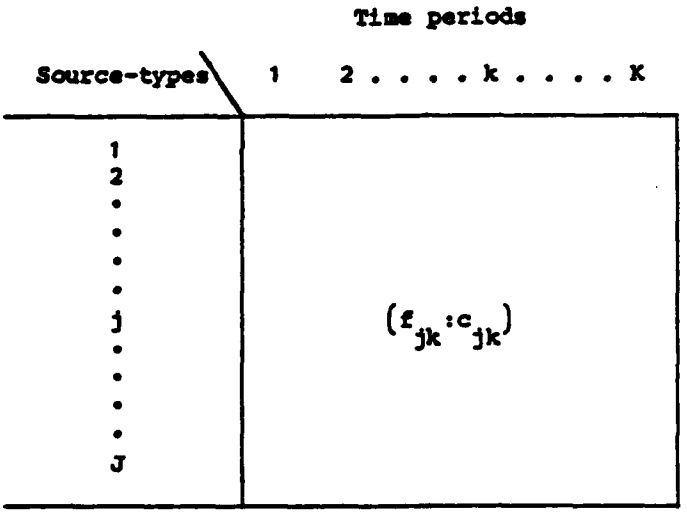


Figure 6. The evidence matrix

since there can be differential source reliability, varying numbers of reports per source, and the sources may not be "independent." While the challenge of understanding how people incorporate such factors into their judgments is formidable, the complexity of inferences in real world settings requires that attention be paid to these issues.

CONCLUSION

In considering the role of ambiguity and uncertainty in inferential judgments, we have developed a quantitative model that accounts for much existing data as well as our own experimental findings. Furthermore, we have shown how this model relates to Keynes' idea of the weight of evidence, the non-additivity of complementary probabilities, risky choice, and current work on cognitive heuristics. Moreover, since inference involves "going beyond the information given" (Bruner, 1957), an important way to do this is to construct, via imagination, "what might have been" or "what might be." Such constructions, whether the result of a cognitive simulation process as proposed here, or more elaborate processes (as in resolving surprise), pose an interesting and important trade-off for the organism. On the one hand, there are costs of investing in imagination; increased mental effort and the discomfort that results from greater uncertainty. On the other hand, the benefits of considering the world as it isn't, protects one from overconfidence and its nonadaptive consequences. Thus, finding the appropriate compromise between "what is" and "what might have been" (or, "what might be"), is central to inferences under ambiguity and uncertainty.

REFERENCES

Becker S. W., & Brownson, F. O. What price ambiguity? Or the role of ambiguity in decision making. Journal of Political Economy, 1964, 72, 62-73.

Birnbaum, M. H. Base rates in Bayesian inference: Signal detection analysis of the cab problem. American Journal of Psychology, 1983, 96, 85-94.

Bishop, Y. M. M., Fienberg, S. E., & Holland, P. W. Discrete multivariate analysis: Theory and practice. Cambridge, MA: MIT Press, 1975.

Bruner, J. S. Going beyond the information given. In J. S. Bruner et al. (Eds.), Contemporary approaches to cognition. Cambridge, MA: Harvard University Press, 1957.

Cohen, L. J. The probable and the provable. Oxford: Clarendon Press, 1977.

Cohen, L. J. Can human irrationality be experimentally demonstrated? The Behavioral and Brain Sciences, 1981, 4, 317-370.

de Finetti, B. Probabilities of probabilities: A real problem or a misunderstanding? In A. Aykac & C. Brumat (Eds.), New directions in the application of Bayesian methods. Amsterdam: North-Holland, 1977.

Edwards, W. The theory of decision making. Psychological Bulletin, 1954, 51, 380-417.

Edwards, W. Conservatism in human information processing. In B. Kleinmuntz (Ed.), Formal representation of human judgment. New York: Wiley, 1968.

Edwards, W., & von Winterfeldt, D. Public disputes about risky technologies: Stakeholders and arenas. Social Science Research Institute, University of Southern California, 1982.

Einhorn, H. J., & Hogarth, R. M. Behavioral decision theory: Processes of judgment and choice. Annual Review of Psychology, 1981, 32, 53-88.

Einhorn, H. J., & Hogarth, R. M. A theory of diagnostic inference: Judging causality. University of Chicago, Graduate School of Business, Center for Decision Research, January 1984.

Ellsberg, D. Risk, ambiguity, and the Savage axioms. Quarterly Journal of Economics, 1961, 75, 643-669.

Fellner, W. Distortion of subjective probabilities as a reaction to uncertainty. Quarterly Journal of Economics, 1961, 75, 670-692.

Fischhoff, B., Slovic, P., & Lichtenstein, S. Subjective sensitivity analysis. Organizational Behavior and Human Performance, 1979, 23, 339-359.

Fishburn, P. C. Ellsberg revisited: A new look at comparative probability. Annals of Statistics, 1983a, 11, 1047-1059

Fishburn, P. C. Normative theories of decision making under risk and under uncertainty. Working paper, presented at Harvard University Conference, "Decision Making: Descriptive, Normative, and Prescriptive Interactions," June 1983b.

Gärdenfors, P., & Sahlin, N.-E. Unreliable probabilities, risk taking, and decision making. Synthese, 1982, 53, 361-386.

Gärdenfors, P., & Sahlin, N.-E. Decision making with unreliable probabilities. British Journal of Mathematical and Statistical Psychology, 1983, 36, 240-251.

Hammond, K. R., Hursch, C. J., & Todd, F. J. Analyzing the components of clinical inference. Psychological Review, 1964, 71, 438-456.

Hammond, K. R., & Summers, D. A. Cognitive control. Psychological Review, 1972, 79, 58-67.

Hershey, J. C., Kunreuther, H. C., & Schoemaker, P. J. H. Sources of bias in assessment procedures for utility functions. Management Science, 1982, 28, 936-954.

Hogarth, R. M., & Kunreuther, H. C. Risk, ambiguity, and insurance. University of Chicago, Graduate School of Business, Center for Decision Research, June 1984.

Jepson, C., Krantz, D., & Nisbett, R. E. Inductive reasoning: Competence or skill? The Behavioral and Brain Sciences, 1983, 6.

Kahneman, D., Slovic, P., & Tversky, A. Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press, 1982.

Kahneman, D., & Tversky, A. Prospect theory: An analysis of decision under risk. Econometrica, 1979, 47, 263-291.

Kahneman, D., & Tversky, A. The simulation heuristic. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press, 1982.

Kamin, L. J. The science and politics of IQ. Potomac, MD: Erlbaum, 1974.

Karmarkar, U. S. Subjectively weighted utility: A descriptive extension of the expected utility model. Organizational Behavior and Human Performance, 1978, 21, 61-72.

Keynes, J. M. A treatise on probability. London: Macmillan, 1921.

Koopman, B. O. The axioms and algebra of intuitive probability. Annals of Mathematics, 1940, 41, 269-292.

Kyburg, H. E., Jr. Rational belief. The Behavioral and Brain Sciences, 1983, 6, 231-273.

Lopes, L. L. Averaging rules and adjustment processes: The role of averaging in inference. Wisconsin Human Information Processing Program, University of Wisconsin, Report 13, 1981.

Lopes, L. L. Some thoughts on the psychological concept of risk. Journal of Experimental Psychology: Human Perception and Performance, 1983, 9, 137-144.

Marschak, J. Personal probabilities of probabilities. Theory and Decision, 1975, 6, 121-153.

Morris, P. Utility, probability, and decision theory. Paper presented at ORSA/TIMS Meeting, April 1983.

Nisbett, R. E., Krantz, D., Jepson, C., & Kunda, Z. The use of statistical heuristics in everyday inductive reasoning. Psychological Review, 1983, 90, 339-363.

Peterson, C. R., & Beach, L. R. Man as an intuitive statistician. Psychological Bulletin, 1967, 68, 29-46.

Rappoport, A., & Wallsten, T. S. Individual decision behavior. Annual Review of Psychology, 1972, 23, 131-176.

Savage, L. J. The foundations of statistics. New York: Wiley, 1954.

Schum, D. A. Current developments in research on cascaded inference processes. In T. S. Wallsten (Ed.), Cognitive processes and decision behavior. Hillsdale, NJ: Erlbaum, 1980.

Schum, D. A., & DuCharme, W. M. Comments on the relationship between the impact and the reliability of evidence. Organizational Behavior and Human Performance, 1971, 6, 111-131.

Schum, D. A., & Kelley, C. W. A problem in cascaded inference: Determining the inferential impact of confirming and conflicting reports from several unreliable sources. Organizational Behavior and Human Performance, 1973, 10, 404-423.

Shafer, G. A. A mathematical theory of evidence. Princeton, NJ: Princeton University Press, 1976.

Shafer, G. A. Two theories of probability. Philosophy of Science Association, 1978, 2, 1-25.

Slovic, P., & Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 6, 649-744.

Slovic, P., Lichtenstein, S., & Fischhoff, B. Behavioral decision theory. Annual Review of Psychology, 1977, 28, 1-39.

Thaler, R. Toward a positive theory of consumer choice. Journal of Economic Behavior and Organization, 1980, 1, 39-60.

Tversky, A., & Kahneman, D. Variants of uncertainty. Cognition, 1982, 11, 143-157.

Wallsten, T. S., Forsyth, B. H., & Budescu, D. Stability and coherence of health experts' upper and lower subjective probabilities about dose-response functions. Organizational Behavior and Human Performance, 1983, 31, 277-302.

Winer, B. J. Statistical principles in experimental design. New York: McGraw Hill, 1963.

Yates, J. F., & Zukowski, L. G. Characterization of ambiguity in decision making. Behavioral Science, 1976, 21, 19-25.

Zadeh, L. A. Fuzzy sets as a basis for a theory of possibility. Fuzzy Sets and Systems, 1978, 1, 3-28.

Zellner, A. An efficient method of estimating seemingly unrelated regressions and tests for aggregation bias. Journal of the American Statistical Association, 1962, 57, 348-368.

FOOTNOTES

This research was supported by a contract from the Office of Naval Research. We wish to thank Haim Mano, Ann McGill, and Diane Vitiello for their assistance on this project. In addition, the following people provided useful comments on earlier versions of the manuscript: Tom Dyckman, William Goldstein, Steven Hoch, Joshua Klayman, Howard Kunreuther, John Payne, Nils-Eric Sahlin, Amos Tversky, and Thomas S. Wallsten.

[1]At $p_A = 0, 1$ there is no ambiguity. Hence, the relation between p_A and $S(p_A)$ should be discontinuous. Indeed, the lack of ambiguity at the end points provides a rationale for the discontinuity of the decision-weight function and this implies the "certainty effect" of prospect theory (i.e., the value of sure gambles is heightened either positively or negatively).

[2]A listing of the program is available from the authors.

APPENDIX A

This appendix considers the effects of different assumptions concerning the weights given in imagination to values of p greater and smaller than p_A. In equation (4), differential weighting is achieved by the β parameter; i.e., $k_g = \theta(1 - p_A)$ and $k_s = \theta p_A^{\beta}$. However, one could also consider linear weighting schemes where the weights given to θp_A and $\theta(1 - p_A)$ sum to one (i.e., a weighted averaging process), or where the weights do not sum to one. For the former, let

$$\begin{aligned} k_g - k_s &= \theta w(1 - p_A) - \theta(1 - w)p_A \\ &= \theta(w - p_A) \end{aligned} \tag{A.1}$$

where $0 \leq w \leq 1$ is the relative weight given to greater values. Substituting (A.1) into equation (6), we obtain,

$$S_1(p_A) = p_A + \theta\,(w - p_A) \tag{A.2}$$

where, $S_1(p_A)$ is used to denote alternative model 1. Note that in this model, $S_1(p_A)$ is regressive with respect to p. Although this model has appealing features, it is easy to show that it does not capture some aspects of our model and data. Specifically, it always predicts additivity of judgments of complementary events, i.e.,

$$\begin{aligned} S_1(p_A) + S_1(1 - p_A) &= p_A + \theta\,(w - p_A) + (1 - p_A) \\ &\quad + \theta\,[(1 - w) - (1 - p_A)] = 1 \end{aligned} \tag{A.3}$$

However, non-additivity will occur if the weights accorded to $\theta(1 - p_A)$ and θp_A do not sum to one. A special case of this model, which we denote $S_2(p_A)$, and which is similar to the $S(p_A)$ model used in the paper, is one where,

$$k_g - k_s = \theta(1 - p_A) - \theta m p_A \quad (m > 0) \qquad (A.4)$$

This yields,

$$S_2(p_A) = p_A + \theta\,[1 - p_A - m p_A] \qquad (A.5)$$

such that the additivity conditions are,

$$S_2(p_A) + S_2(1 - p_A) = p_A + \theta\,[1 - p_A - m p_A] + (1 - p_A) + \theta\,[p_A - m(1-p_A)]$$

$$= 1 + \theta\,(1-m) \qquad (A.6)$$

Thus, for $m > 1$, the model predicts sub-additivity; for $m = 1$, additivity; and for $m < 1$, super-additivity. The difference between $S_2(p_A)$ and $S(p_A)$ is that the former predicts a constant amount of non-additivity irrespective of the value of p_A. In the $S(p_A)$ model, the level of p_A affects the amount of additivity. This is shown in equation (7), which is reproduced here for convenience,

$$S(p_A) + S(1 - p_A) = 1 + \theta\,[1 - p_A^{\beta} - (1 - p_A)^{\beta}] \qquad (A.7)$$

OFFICE OF NAVAL RESEARCH

Engineering Psychology Program

TECHNICAL REPORTS DISTRIBUTION LIST

OSD

CAPT Paul R. Chatelier
Office of the Deputy Under Secretary
of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D. C. 20301

Dr. Dennis Leedom
Office of the Deputy Under Secretary
of Defense (C^3I)
Pentagon
Washington, D. C. 20301

Department of the Navy

Engineering Psychology Group
Office of Naval Research
Code 442EP
800 N. Quincy St.
Arlington, VA 22217 (3 cys.)

Aviation & Aerospace Technology
Programs
Code 210
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR. Paul E. Girard
Code 252
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Physiology Program
Office of Naval Research
Code 441NP
800 North Quincy Street
Arlington, VA 22217

Dr. Edward H. Huff
Man-Vehicle Systems Research Division
NASA Ames Research Center
Moffett Field, CA 94035

Department of the Navy

Dr. Andrew Rechnitzer
Office of the Chief of Naval
Operations, OP952F
Naval Oceanography Division
Washington, D.C. 20350

Manpower, Personnel & Training
Programs
Code 270
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Mathematics Group
Code 411-MA
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Statistics and Probability Group
Code 411-S&P
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Information Sciences Division
Code 433
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR Kent S. Hull
Helicopter/VTOL Human Factors Office
NASA-Ames Research Center MS 239-21
Moffett Field, CA 94035

Dr. Carl E. Englund
Naval Health Research Center
Environmental Physiology
P.O. Box 85122
San Diego, CA 92138

Department of the Navy

Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Mr. R. Lawson
ONR Detachment
1030 East Green Street
Pasadena, CA 91106

CDR James Offutt
Office of the Secretary of Defense
Strategic Defense Initiative Organization
Washington, D.C. 20301-7100

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375

Dr. Michael Melich
Communications Sciences Division
Code 7500
NaVAL Research Laboratory
Washington, D.C. 23075

Dr. J. S. Lawson
Naval Electronic Systems Command
NELEX-06T
Washington, D. C. 20360

Dr. Neil McAlister
Office of Chief of Naval Operations
Command and Control
OP-094H
Washington, D. C. 20350

Naval Training Equipment Center
ATTN: Technical Library
Orlando, FL 32813
Naval Research Laboratory
Washington, D.C. 20375

Department of the Navy

Dr. Robert G. Smith
Office of the Chief of Naval
Operations, OP987H
Personnel Logistics Plans
Washington, D. C. 20350

Combat Control Systems Department
Code 35
Naval Underwater Systems Center
Newport, RI 02840

Human Factors Department
Code N-71
Naval Training Equipment Center
Orlando, FL 32813

Dr. Alfred F. Smode
Training Analysis and Evalustion
Group
Naval Training & Equipment Center
Orlando, FL 32813

Human Factors Engineering
Code 8231
Naval Ocean Systems Center
San Diego, CA 92152

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, CA 93940

Mr. H. Talkington
Engineering & Computer Science
Code 09
Naval Ocean Systems Center
San Diego, CA 92152

Department of the Navy

Mr. Paul Heckman
Naval Ocean Systems Center
San Diego, CA 92152

Dr. Ross Pepper
Naval Ocean Systems Center
Hawaii Laboratory
P. O. Box 997
Kailua, HI 96734

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D. C. 20380

Dr. L. Chmura
Naval Research Laboratory
Code 7592
Computer Sciences & Systems
Washington, D. C. 20375

Office of the Chief of Naval
Operations (OP-115)
Washington, D.C. 20350

Professor Douglas E. Hunter
Defense Intelligence College
Washington, D.C. 20374

CDR C. Hutchins
Code 55
Naval Postgraduate School
Monterey, CA 93940

Human Factors Technology Administrator
Office of Naval Technology
Code MAT 0722
800 N. Quincy Street
Arlington, VA 22217

CDR Tom Jones
Naval Air Systems Command
Human Factors Programs
NAVAIR 330J
Washington, D. C. 20361

Department of the Navy

Commander
Naval Air Systems Command
Crew Station Design
NAVAIR 5313
Washington, D. C. 20361

Mr. Philip Andrews
Naval Sea Systems Command
NAVSEA 61R
Washington, D. C. 20362

Commander
Naval Electronics Systems Command
Human Factors Engineering Branch
Code 81323
Washington, D. C. 20360

Mr. Herb Marks
Naval Surface Weapons Center
NSWC/DL
Code N-32
Dahlgren, VA 22448

Mr. Milon Essoglou
Naval Facilities Engineering Command
R&D Plans and Programs
Code 03T
Hoffman Building II
Alexandria, VA 22332

CAPT Robert Biersner
Naval Biodynamics Laboratory
Michoud Station
Box 29407
New Orleans, LA 70189

Dr. Arthur Bachrach
Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, MD 20014

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Department of the Navy

Head
Aerospace Psychology Department
Code L5
Naval Aerospace Medical Research Lab
Pensacola, FL 32508

Commanding Officer
Naval Health Research Center
San Diego, CA 92152

Dr. Jerry Tobias
Auditory Research Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Navy Personnel Research and
Development Center
Planning & Appraisal Division
San Diego, CA 92152

Dr. Robert Blanchard
Navy Personnel Research and
Development Center
Command and Support Systems
San Diego, CA 92152

CDR J. Funaro
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Stephen Merriman
Human Factors Engineering Division
Naval Air Development Center
Warminster, PA 18974

Mr. Jeffrey Grossman
Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

Human Factors Engineering Branch
Code 4023
Pacific Missile Test Center
Point Mugu, CA 93042

Department of the Navy

Dean of the Academic Departments
U. S. Naval Academy
Annapolis, MD 21402

Dr. W. Moroney
Naval Air Development Center
Code 602
Warminster, PA 18974

Human Factor Engineering Branch
Naval Ship Research and Development
Center, Annapolis Division
Annapolis, MD 21402

Dr. Harry Crisp
Code N 51
Combat Systems Department
Naval Surface Weapons Center
Dahlgren, VA 22448

Mr. John Quirk
Naval Coastal Systems Laboratory
Code 712
Panama City, FL 32401

Department of the Army

Dr. Edgar M. Johnson
Technical Director
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Technical Director
U. S. Army Human Engineering Labs
Aberdeen Proving Ground, MD 21005

Director, Organizations and
Systems Research Laboratory
U. S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Mr. J. Barber
HQS, Department of the Army
DAPE-MBR
Washington, D.C. 20310

Department of the Air Force

Dr. Kenneth R. Boff
AF AMRL/HE
Wright-Patterson AFB, OH 45433

U.S. Air Force Office of
Scientific Research
Life Science Directorate, NL
Bolling Air Force Base
Washington, D.C. 20332

AFHRL/LRS TDC
Attn: Susan Ewing
Wright-Patterson AFB, OH 45433

Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Dr. Earl Alluisi
Chief Scientist
AFHRL/CCN
Brooks Air Force Base, TX 78235

Dr. R. K. Dismukes
Associate Director for Life Sciences
AFOSR
Bolling AFB
Washington, D.C. 20332

Foreign Addresses

Dr. Kenneth Gardner
Applied Psychology Unit
Admiralty Marine Tech. Estab.
Teddington, Middlesex TW11 OLN
England

Human Factors
P.O. Box 1085
Station B
Rexdale, Ontario
Canada M9V 2B3

Foreign Addresses

Dr. A. D. Baddeley
Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF England

Other Government Agencies

Defense Technical Information Center
Cameron Station, Bldg. 5
Alexandria, VA 22314 (12 copies)

Dr. Clinton Kelly
Defense Advanced Research Projects
Agency
1400 Wilson Blvd.
Arlington, VA 22209

Dr. M. C. Montemerlo
Human Factors & Simulation
Technology, RTE-6
NASA HQS
Washington, D.C. 20546

Other Organizations

Ms. Denise Benel
Essex Corporation
333 N. Fairfax Street
Alexandria, VA 22314

Dr. Andrew P. Sage
First American Prof. of Info. Tech.
Assoc. V.P. for Academic Affairs
George Mason University
4400 University Drive
Fairfax, VA 22030

Other Organizations

Dr. Robert R. Mackie
Human Factors Research Division
Canyon Research Group
5775 Dawson Avenue
Goleta, CA 93017

Dr. Amos Tversky
Dept. of Psychology
Stanford University
Stanford, CA 94305

Dr. H. McI. Parsons
Essex Corporation
333 N. Fairfax St.
Alexandria, VA 22314

Dr. Jesse Orlansky
Institute for Defense Analyses
1801 N. Beauregard Street
Alexandria, VA 22043

Dr. J. O. Chinnis, Jr.
Decision Science Consortium, Inc.
7700 Leesburg Pike
Suite 421
Falls Church, VA 22043

Dr. T. B. Sheridan
Dept. of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. Paul E. Lehner
PAR Technology Corp.
Seneca Plaza, Route 5
New Hartford, NY 13413

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Other Organizations

Dr. Harry Snyder
Dept. of Industrial Engineering
Virginia Polytechnic Institute
and State University
Blacksburg, VA 24061

Dr. Stanley Deutsch
NAS-National Research Council (COHF)
2101 Constitution Avenue, N.W.
Washington, D.C. 20418

Dr. Amos Freedy
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. Robert Fox
Dept. of Psychology
Vanderbilt University
Nashville, TN 37240

Dr. Meredith P. Crawford
American Psychological Association
Office of Educational Affairs
1200 17th Street, N.W.
Washington, D.C. 20036

Dr. Deborah Boehm-Davis
Dept. of Psychology
George Mason University
4400 University Drive
Fairfax, VA 22030

Dr. Howard E. Clark
NAS-NRC
Commission on Engrg. & Tech. Systems
2101 Constitution Ave., N.W.
Washington, D.C. 20418

Other Organizations

Dr. Charles Gettys
Department of Psychology
University of Oklahoma
455 West Lindsey
Norman, OK 73069

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Boulder, CO 80309

Dr. James H. Howard, Jr.
Department of Psychology
Catholic University
Washington, D. C. 20064

Dr. William Howell
Department of Psychology
Rice University
Houston, TX 77001

Dr. Christopher Wickens
Department of Psychology
University of Illinois
Urbana, IL 61801

Mr. Edward M. Connelly
Performance Measurement
Associates, Inc.
1909 Hull Road
Vienna, VA 22180

Professor Michael Athans
Room 35-406
Massachusetts Institute of
Technology
Cambridge, MA 02139

Dr. Edward R. Jones
Chief, Human Factors Engineering
McDonnell-Douglas Astronautics Co.
St. Louis Division
Box 516
St. Louis, MO 63166

Other Organizations

Dr. Babur M. Pulat
Department of Industrial Engineering
North Carolina A&T State University
Greensboro, NC 27411

Dr. Lola Lopes
Information Sciences Division
Department of Psychology
University of Wisconsin
Madison, WI 53706

National Security Agency
ATTN: N-32, Marie Goldberg
9800 Savage Road
Ft. Meade, MD 20722

Dr. Stanley N. Roscoe
New Mexico State University
Box 5095
Las Cruces, NM 88003

Mr. Joseph G. Wohl
Alphatech, Inc.
3 New England Executive Park
Burlington, MA 01803

Dr. Marvin Cohen
Decision Science Consortium, Inc.
Suite 721
7700 Leesburg Pike
Falls Church, VA 22043

Dr. Robert Wherry
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

Dr. William R. Uttal
Institute for Social Research
University of Michigan
Ann Arbor, MI 48109

Dr. William B. Rouse
School of Industrial and Systems
Engineering
Georgia Institute of Technology
Atlanta, GA 30332

Other Organizations

Dr. Richard Pew
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Dr. Douglas Towne
University of Southern California
Behavioral Technology Lab
3716 S. Hope Street
Los Angeles, CA 90007

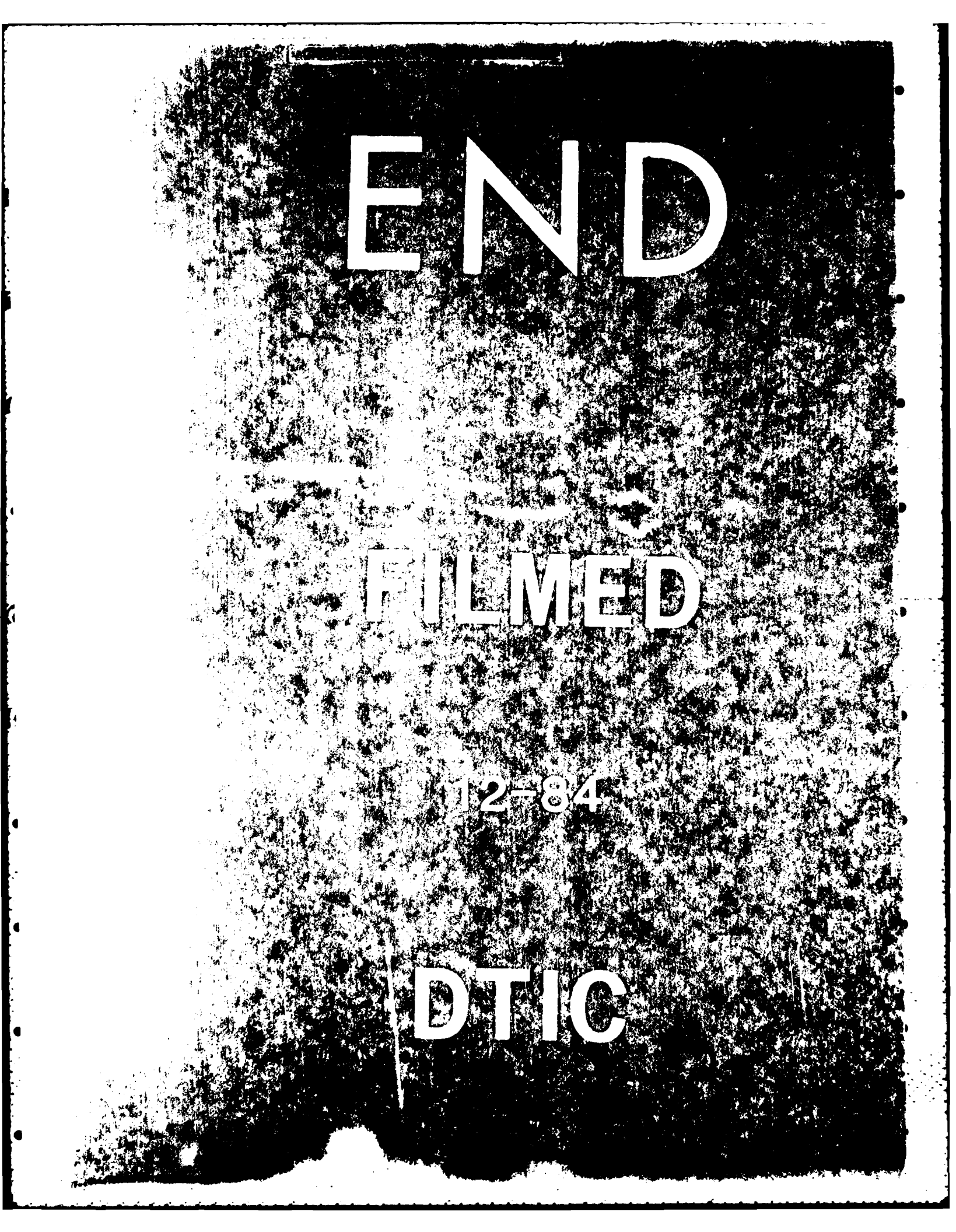
Dr. David J. Getty
Bolt Beranek & Newman, Inc.
50 Moulton street
Cambridge, MA 02238

Dr. John Payne
Graduate School of Business
Administration
Duke University
Durham, NC 27706

Dr. Baruch Fischhoff
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Alan Morse
Intelligent Software Systems Inc.
160 Old Farm Road
Amherst, MA 01002

Dr. J. Miller
Florida Institute of Oceanography
University of South Florida
St. Petersburg, FL 33701


END
FILMED
12-84
DTIC